

UNIVERSITY OF ECONOMICS, PRAGUE
DEPARTMENT OF BANKING AND INSURANCE
FACULTY OF FINANCE



Pricing and modeling credit risk

A thesis presented for the degree of
Doctor of Philosophy

Author:
Marek KOLMAN

Supervisor:
Prof. RNDr. Jiří WITZANY, Ph.D.

February 2017

Pricing and modeling credit risk

Marek Kolman

Abstract

The thesis covers a wide range of topics from the credit risk modeling with the emphasis put on pricing of the claims subject to the default risk. Starting with a separate general contingent claim pricing framework the key topics are classified into three fundamental parts: firm-value models, reduced-form models, portfolio problems, with a possible finer sub-classification. Every part provides a theoretical discussion, proposal of self-developed methodologies and related applications that are designed so as to be close to the real-world problems.

The text also reveals several new findings from various fields of credit risk modeling. In particular, it is shown (i) that the stock option market is a good source of credit information, (ii) how the reduced-form modeling framework can be extended to capture more complicated problems, (iii) that the double t copula together with a self-developed portfolio modeling framework outperforms the classical Gaussian copula approaches. Many other, partial findings are presented in the relevant chapters and some other results are also discussed in the Appendix.

Declaration

I hereby declare that all sources are properly acknowledged in the text and all errors are my own responsibility.

*In Prague,
February 4, 2017*

.....
Marek Kolman

Acknowledgements

I am forever indebted to my supervisor Professor Jiří Witzany for his support, enriching suggestions, permanent push forward and infinite patience. Special thanks also belong to my family and in particular to my father Lubomír who passed away before I was able to finish this thesis. My thanks also belong to J&T Banka and my former colleague Radoslav Míšek for creating such working conditions that allowed me to write the majority of this text.

Contents

Introduction	4
I Arbitrage pricing theory	6
I.1 Overview of the main results	6
I.1.1 Some important results from the stochastic calculus	6
I.1.2 Cross variation and quadratic variation	10
I.1.3 Partial differential equations	11
I.1.4 Self-financing trading strategies and arbitrage	14
I.1.5 Martingale measure and Girsanov theorem	15
I.1.6 Pricing of contingent claims and the change of numéraire	18
I.1.7 Practical observations and comments	21
II Firm-value models	23
II.1 Firm-value approach to credit risk	23
II.1.1 Preliminaries and the general modeling environment	24
II.1.2 General risk-neutral valuation formula	26
II.1.3 PDE framework for credit risk pricing	30
II.1.4 Models with closed-forms	35
II.1.4.1 Merton's model	35
II.1.4.2 Merton's model with stochastic r (Hull-White)	39
II.1.4.3 Zhou's jump diffusion model	41
II.1.4.4 The Hestonian approach	43
II.1.5 The analysis	46
II.1.5.1 Proposed general method to the model calibration	47
II.1.5.2 The data	49
II.1.5.3 The results	51
II.1.5.4 Conclusion	56
III Reduced-form models	58
III.1 Definitions and essential concepts	59
III.1.1 Hazard function and cumulative hazard function	60
III.1.2 Hazard process and cumulative hazard process	62
III.2 The fundamental pricing equation	63
III.2.1 Applications of the fundamental PDE	71
III.2.1.1 Pricing of a corporate bond and its embedded option	71
III.2.1.2 Pricing the counterparty risk in an IRS contract	75
III.2.1.3 Calibration, pricing and valuation of CDS	80

III.2.1.4	Pricing of CDS options	84
III.2.2	Extensions	89
III.2.2.1	Counterparty risk and extension for correlated defaults	89
III.2.2.2	Multidimensional jump–diffusion model and PIDE	96
III.3	A defaultable binomial tree	103
III.3.1	The model setup	103
III.3.2	Approximating extension to coupon-bearing bonds	110
III.3.3	European options on defaultable bonds	111
III.3.4	American options	113
III.3.5	Revisiting barrier options on a stock. The binomial tree approach	113
III.3.6	Barrier options	115
III.3.7	Applications	119
III.3.8	Numerical validation against Monte Carlo and PDE	123
IV	Portfolio problems	127
IV.1	Synthetic CDO modeling	127
IV.1.1	The current CDO market practice and its imperfections	128
IV.1.2	The mechanics of a CDO deal	129
IV.1.3	General pricing framework	130
IV.1.4	The terms $\mathcal{A}, \mathcal{B}, \mathcal{C}$ in a bottom–up framework	131
IV.1.5	The expected tranche principal	133
IV.1.6	$Q(t x)$ and the factor function	136
IV.1.7	The Normal Inverse Gaussian (NIG) copula	137
IV.1.8	The double t copula	140
IV.1.9	The stochastic correlation model	143
IV.1.10	$Q(t x)$ and the Archimedean copulas	145
IV.1.11	The data, calibration procedure and assessment of the models	147
IV.1.12	Concluding remarks	150
IV.2	Modeling the portfolio risk	152
IV.2.1	Generator of default rates	154
IV.2.2	Generating the distribution of number of defaults	155
IV.2.3	Mapping of default losses to the default events	156
IV.2.4	Deterministic and stochastic recoveries	156
IV.2.5	The double t specification	158
IV.2.6	The convergence to the Vasicek’s LHP	159
IV.2.7	Speed issues and recurrence relations	160
IV.2.8	Dataset and benchmarking	165
IV.2.9	Concluding remarks	169
Conclusion		171

A	The fixed-income toolkit	172
A.1	Basic continuous-time processes	172
A.2	Fundamentals of the HJM framework	173
A.3	Caps, floors, swaptions and Black formulas	174
A.4	Caps, floors and swaptions in the short-rate models	177
B	Reduced Convolution	180
C	Numerical solutions to PDEs	182
C.1	One-dimensional PDEs	183
C.2	Two-dimensional PDEs	186
D	Estimation of SDE/JDE parameters using PDE/PIDE	192
E	Proof of the Black-Scholes model with Hull-White's r	194

Introduction

Credit risk has always been the major risk in the banking business. If the price of an investment such as stock falls, such an investment still might be sold even though with a loss incurred. If, however, a counterparty fails this might be a critical event for more players as their businesses with the defaulted counterparty freeze. Also, when the default is large, a contagion looms over the whole market and the most players resort to waiting instead of doing business. The recent past has proved this to be very true. The ability to correctly measure the credit risk is therefore critical for any financial institution. An even more challenging discipline is the pricing of the credit risk as one can not only rely on some statistical estimations but also the expectations of the market participants shall be taken into account.

Motivated by these simple facts the aim of this thesis is to provide a detailed insight into the techniques of modeling credit risk with the focus on pricing of securities subject to credit risk while some of the techniques are self-developed to better fit the problem. There are, in general, many possibilities or perspectives how this 'credit risk modeling' can be accomplished. In this text, we try to apply methods that we believe are sufficiently general and enough versatile for various real-world pricing cases. That is, we discuss methods that have some application potential, yield reproducible results and are not far away from the usual market practice. We also desire to deliver a reasonable balance between the theory and the applications so a large portion of the theoretical exposition is supported with applications. Our intention is also to provide a self-contained text such that the audience does not need to jump to some other literature but all the theory is provided here.

This thesis is structured into four master sections, referred to as the *Parts*. The Part **I** provides a general discussion targeting the main results from the stochastic analysis and it has no particular focus on the credit risk. It is a collection of results from various recognized authors yet we removed what we believe is not necessary for the applications but rather for the theory itself. This is merely an entry section for a better understanding of what follows later in the text. We try to provide a succinct, condensed, 15 pages chapter without leaving aside anything important that could be missing later in the text. Arguably the most cited passage of this part is the section discussing the linkages of the partial differential equations, expectations and stochastic differential equations.

The Parts **II**, **III** are dedicated to the firm-value and the reduced-form credit models which represent two main credit risk modeling methodologies. Although perhaps more space is allocated for the reduced-form models in the Part **III**, we find it very useful to include the firm-value models in the Part **II** as well. We have found limited literature which would be sufficiently rigorous, detailed and would cover *both* these model classes so that some sort of comparison between the two is possible and this is what we have done. We reckon this makes the text much more informative as the reader is always informed about the 'alternative' and the differences between both approaches.

The portfolio modeling is covered by the Part **IV**. It adopts some features from both the firm-value modeling and the reduced-form modeling and therefore it is naturally presented in the end. We took a journey into the CDO modeling in order to find the best

possible model for modeling the CDO prices as there are many choices. Consistently with this modeling approach we present a self-developed framework for portfolio credit risk modeling which captures some of the current market phenomena such as tail correlations, stochastic recoveries and more. All these methodologies are supplemented with proposed computational techniques to leverage the benefits.

Part I

Arbitrage pricing theory

I.1 Overview of the main results

The main objective of this compressed Part I is to present necessary results for the valuation of contingent claims *free of any assumptions about the credit risk*. Inclusion of this section can be motivated by the fact that later in the text (particularly Part II and Part III) we deal with contingent claims that are subject to stochastic processes. We thus find it necessary to present the foundations of contingent claims pricing at the entry point of this thesis and also it helps the reader to familiarize himself with the notation used in the entire text.

This section is organized in the following way. We firstly summarize some essential results from the stochastic analysis, then we link this theory to the finance-related topics such as the risk-neutral pricing. The end of this part is dedicated to practical comments based on the experience the author has collected. In this entry part we mostly follow the exposition provided by (Björk, 2009), (Shreve, 2004), (Andersen and Piterbarg, 2010) and more rigorous (Musiela and Rutkowski, 2010).

I.1.1 Some important results from the stochastic calculus

Since the processes that underlie the pricing problems are assumed to occur in the continuous times, we skip the discrete-time analysis and move directly to the continuous-time stochastic calculus. A natural start is from the analysis of Itô's processes.

Consider a d -dimensional Itô's process

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t), \quad (\text{I.1.1})$$

where $\mu : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, $\sigma : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^p$ and W is a p -dimensional standard Wiener process. This is the basic general setup. Also define

$$|\sigma(t, X(t))| = \sqrt{\sum_{i=1}^d \sum_{j=1}^p \left| \sigma(t, X(t))_{i,j} \right|^2} = \sqrt{\text{trace} \left(\sigma(t, X(t)) \sigma(t, X(t))^T \right)}. \quad (\text{I.1.2})$$

Here, we note that $|\cdot|$ is a standard Euclidean norm (also known as the Frobenius norm in the case of matrix). In order to ensure some desirable properties of Itô's processes, it is necessary to introduce L^1, L^2, H^2 spaces.

Definition I.1.1 (L^1, L^2 space)

Assume a process X such as (I.1.1). Then we say μ is in L^1 if

$$\int_0^t |\mu(s, X(s))| ds < \infty.$$

If

$$\int_0^t |\sigma(s, X(s))|^2 ds < \infty,$$

then σ belongs to L^2 .

Throughout this text, we assume that both μ and σ adapted and are in L^1 and L^2 , respectively.

Definition I.1.2 (H^2 space)

Consider $|\sigma(t, X(t))|^2$ as in (I.1.2). We say, σ belongs to H^2 , if for all $t \in [0, T]$ we have

$$\mathbb{E} \left[\int_0^t |\sigma(s, X(s))|^2 ds \right] < \infty.$$

This square-integrability condition is useful for justification of Itô's isometry, which is introduced by means of the Definition I.1.5. Because the valuation often requires to work with martingales, we present a formal definition.

Definition I.1.3 (Martingale)

Let M be an adapted (possibly vector-valued) process, with a finite expectation $\mathbb{E}[|M(t)|] < \infty$ for $t \in [0, T]$. Then M is a martingale (under \mathbb{P}) if for all $s, t \in [0, T], t \leq s$

$$\mathbb{E}[M(s) | \mathcal{F}(t)] = M(t).$$

Replacing the equality sign with \leq or \geq leads to definition of a supermartingale or a submartingale, respectively.

The concept of martingales is the base of derivatives pricing. Note that in the definition we used \mathbb{P} to denote a measure. This is a general label of a measure but later in the text \mathbb{P} will denote an empirical measure under which the processes are, in general, not martingales.

Definition I.1.4 (Semimartingale)

Let M be a martingale and let A is a process of finite variation.¹ Then $X(t) = M(t) + A(t)$ is semimartingale under \mathbb{P} for all $t \in [0, T]$.

The semimartingale definition says that a (semimartingale) process X can be decomposed into a martingale M and a 'simpler' process A . Note that Wiener-driven process are of infinite (total, i.e. first) variation hence A can not be a Wiener-driven process (but M can).

Let us also define the Itô's integral and list several useful properties of it.

¹a finite variation (or bounded variation) process is a process X whose total variation $\text{TV}[X](T) = \lim_{\|\Pi\| \rightarrow 0} \sum_i |X(t_{i+1}) - X(t_i)|$, where $\|\Pi\|$ is the partition of the mesh, is finite (bounded). This rules out the Wiener process which has infinite variation and allows for cadlag processes, which in particular means all continuous differentiable processes and also e.g. the Poisson process.

Definition I.1.5 (Itô integral and its properties)

Under the assumption that σ is in H^2 as postulated by the Definition I.1.2, let

$$I(t) = \int_0^t \sigma(s, X(s)) dW(s) = \begin{pmatrix} I_1(t) \\ \vdots \\ I_d(t) \end{pmatrix},$$

denote the Itô (stochastic) integral (which can be a vector). It has the following properties:

- (i) I is adapted to the filtration $\mathbb{F} = \{\mathcal{F}(t)\}_{t \geq 0}$ ($I(t)$ is $\mathcal{F}(t)$ -measurable).
- (ii) $I(t)$ is a (continuous) martingale. $\mathbb{E}[I(t)] = 0$. For a vector $I(t)$ this holds for every its element.
- (iii) Itô isometry: $\mathbb{E}[|I(t)|^2] = \mathbb{E}\left[\int_0^t |\sigma(s, X(s))|^2 ds\right] < \infty$.
- (iv) Covariance

$$\begin{aligned} \mathbb{E}[I(t)I(u)^T] &= \begin{bmatrix} \mathbb{E}[I_1(t)I_1(u)] & \mathbb{E}[I_1(t)I_2(u)] & \cdots & \mathbb{E}[I_1(t)I_d(u)] \\ \mathbb{E}[I_2(t)I_1(u)] & \mathbb{E}[I_2(t)I_2(u)] & \cdots & \mathbb{E}[I_2(t)I_d(u)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}[I_d(t)I_1(u)] & \mathbb{E}[I_d(t)I_2(u)] & \cdots & \mathbb{E}[I_d(t)I_d(u)] \end{bmatrix} \\ &= \mathbb{E}\left[\int_0^{\min[t,u]} \sigma(s, X(s))\sigma(s, X(s))^T ds\right]. \end{aligned}$$

We note that in the one-dimensional case (σ, W, I are one-dimensional components) every element is trivially a scalar.

The Definition I.1.5 requires a comment. The property of any martingale M that

$$\mathbb{E}[|M(t)|^2] < \infty,$$

labels this martingale M as *square-integrable*. Therefore, according to the point (iii), every stochastic integral I is also square-integrable due to the inequality constraint ($< \infty$). However, the square-integrability property is valid if and only if σ is in H^2 . When σ is not in H^2 then we are working 'only' with *local martingales* not martingales. We note that every martingale is also a local martingale but not every local martingale is necessarily a martingale. Throughout this text, we disregard this a rather technical difference (hence we assume σ is in H^2) and we consider all local martingales to be (full) martingales, when not stated otherwise.

For practical applications, it would also be useful to have a tool that would guarantee the existence of σ given a martingale process.

Theorem I.1.1 (Martingale representation theorem)

Let M be a Wiener-driven (local) martingale process. Then it always has the representation

$$dM(t) = \sigma(t, M(t))dW(t).$$

If, to the addition, M is square-integrable, such that M is a true martingale, then σ is in H^2 .

Power of this surprisingly short theorem resides in the property that given a measure under which a process adapted to a Wiener filtration is a martingale implies the above form of the process. For example, consider that we perform the change of measure such that under the new measure the process M is a martingale. In such a case, we do not need to find the differential of M analytically because we are automatically *guaranteed* the form above. The technical remark about the square-integrability is actually just rephrasing of what was already pointed out in connection with the point (iii) of the Definition I.1.5.

Often, we are challenged to work with functions of Itô's processes and for that the following famous result is necessary.

Theorem I.1.2 (Itô's lemma)

Let $f(t) = f(t, X(t)) \in C^{1,2,\dots,2}$ be a function of time and Itô's d -dimensional process X driven by a p -Wiener as defined in (I.1.1). Then in the simplest possible representation, f has the stochastic differential

$$df(t) = \frac{\partial f}{\partial t} dt + \sum_{i=1}^d \frac{\partial f}{\partial X_i(t)} dX_i(t) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 f}{\partial X_i(t) \partial X_j(t)} dX_i(t) dX_j(t).$$

In a more elaborate version, this is

$$\begin{aligned} df(t) &= \left(\frac{\partial f}{\partial t} + \sum_{i=1}^d \mu_i(t, X(t)) \frac{\partial f}{\partial X_i(t)} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 f}{\partial X_i(t) \partial X_j(t)} [\sigma(t, X(t)) \sigma(t, X(t))^T]_{i,j} \right) dt \\ &+ \sum_{i=1}^d \left[\frac{\partial f}{\partial X_i(t)} \left(\sum_{j=1}^p \sigma_{i,j}(t, X(t)) dW_j(t) \right) \right]. \end{aligned}$$

In the theorem above it is assumed that f is evaluated at $(t, X(t))$. The most frequent application of Itô's lemma is for the one-dimensional processes. We provide the one-dimensional version for a quick reference.

Corollary I.1.1 (Itô's lemma in one dimension)

Let $f(t) = f(t, X(t)) \in C^{1,2}$ be a function of time and one-dimensional process X driven by a one-dimensional Wiener process W . Then the stochastic differential of f equals

$$df(t) = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial X(t)} dX(t) + \frac{1}{2} \frac{\partial^2 f}{\partial X(t)^2} [dX(t)]^2.$$

Another very common application of Itô's lemma is then finding the differential of a product of two Itô's processes.

Corollary I.1.2 (Itô's product rule)

Let $f(t) = f(t, X(t), Y(t)) = X(t)Y(t)$, where X and Y are one-dimensional Itô's processes. Then

$$\begin{aligned} df(t) &= \frac{\partial f}{\partial X(t)} dX(t) + \frac{\partial f}{\partial Y(t)} dY(t) \\ &+ \frac{1}{2} \left(\frac{\partial^2 f}{\partial X(t)^2} [dX(t)]^2 + 2 \frac{\partial^2 f}{\partial X(t) \partial Y(t)} dX(t) dY(t) + \frac{\partial^2 f}{\partial Y(t)^2} [dY(t)]^2 \right) \\ &= Y(t) dX(t) + X(t) dY(t) + dX(t) dY(t). \end{aligned} \tag{I.1.3}$$

The Itô's product rule is a particular case of the multidimensional case (Theorem I.1.2). Note that in this particular case $\frac{\partial f}{\partial X(t)} = Y(t)$, $\frac{\partial f}{\partial X(t)} = Y(t)$, $\frac{\partial^2 f}{\partial X(t)^2} = 0$, $\frac{\partial^2 f}{\partial Y(t)^2} = 0$, and $\frac{\partial^2 f}{\partial X(t)\partial Y(t)} = 1$, so the result takes a very simple form. Notice that if X and Y are uncorrelated² processes, then the term $dX(t)dY(t)$ term is 0.

I.1.2 Cross variation and quadratic variation

When working with the stochastic processes, we need the definition of the cross variation and the quadratic variation. The main motivation for introducing these mathematical concepts is that operations with random processes often involve 'multiplication of differentials' of stochastic processes. The product of differentials actually can equivalently be expressed in terms of either *cross variation* or *quadratic variation* (if we analyze the square of a given differential). The product of differentials and the cross/quadratic variation are essentially the same concepts.

For both the cross variation and the quadratic variation we need to introduce a *mesh* Π being a set of time points $\Pi = \{t_0 = 0, t_1, \dots, t_n = T\}$ and a *partition* $\|\cdot\|$ of the mesh denoting the longest subinterval of Π , thus $\|\Pi\| = \max_{i \geq 0} [t_{i+1} - t_i]$. Let us define the cross variation first.

Definition I.1.6 (Cross variation)

Consider two one-dimensional processes X, Y . Then, their cross variation from 0 to T is

$$[X, Y](T) = \lim_{\|\Pi\| \rightarrow 0} \sum_i [(X(t_{i+1}) - X(t_i))(Y(t_{i+1}) - Y(t_i))] = \int_0^T dX(t)dY(t).$$

Here, the integral involving two differentials represents the limiting sum of the differences of X and Y as $\|\Pi\| \rightarrow 0$. Equivalently, it is easy to see that the cross variation can be written in a differential form as

$$d[X, Y](t) = dX(t)dY(t),$$

that is the increment of the cross variation of X, Y is equal to the product of the differentials of X, Y .

Similarly, it is quite easy to define a *quadratic variation*, which is just a cross variation of a process with itself.

Definition I.1.7 (Quadratic variation)

Quadratic variation of a one-dimensional process X is defined as

$$[X, X](T) = [X](T) = \lim_{\|\Pi\| \rightarrow 0} \sum_i (X(t_{i+1}) - X(t_i))^2 = \int_0^T dX(t)dX(t).$$

Equivalently, in a differential and maybe more frequently used notation this is

$$d[X, X](t) = dX(t)dX(t).$$

²the meaning of uncorrelated processes is that $dX(t)dY(t) = 0 \times dt$.

Lemma I.1.1 (Cross variation and quadratic variation of Wiener processes)

The cross variation of two Wiener processes is ρdt . The quadratic variation of a Wiener process from time 0 to T is $W(T)$.

This can be equivalently written as

$$d[W_1, W_2](t) = \rho dt,$$

for the cross variation, and

$$d[W, W](t) = dt,$$

for the quadratic variation, respectively. We note that smooth integrators imply zero quadratic variation.³ Finally note that $[X, Y](t) \neq X(t)Y(t)$.

I.1.3 Partial differential equations

Partial differential equations (PDEs) are very strictly linked to Itô's processes that in the most cases underlie the pricing problems. The main motivation for the usage of PDEs in finance is that they allow the pricing problems to be solved directly using numerical methods, i.e. without the necessity to find analytical solutions which in more general cases might not even be available. If a pricing problem involves several stochastic factors or/and a barrier,⁴ then PDEs provide a convenient and also a relatively general problem-solving framework. In relation with PDEs we highlight particularly the role of Kolmogorov Backward Equation (KBE) and its extension Feynman-Kac theorem which provides a quick tool for construction of PDEs from the Itô's processes and appropriate expectations formed over these Itô's processes. Once the PDE is established, a suitable choice of a numerical technique such as finite differences and a scheme needs to be selected in order to find an exact numerical solution to the pricing problem. Numerical solutions to PDEs are discussed in the Appendix C. A specific (ADI⁵) method for the two-dimensional problems is also provided in (Kolman, 2015b).

The purpose of this section is to introduce the essential tools for formulations of pricing problems that are subject to Itô's processes. We start with the Kolmogorov Backward Equation, then proceed with the Kolmogorov Forward Equation (KFE) and finally we introduce an extensively used theorem in finance, the Feynman-Kac theorem. Throughout this section we assume that some general multidimensional Itô's process X such as (I.1.1) is given.

Let us fix a time frame $[0, T]$, with T being the time of the terminal state and let us assume $u = u(t, X(t))$ represents a conditional expectation of a 'system' which exhibits some *nonrandom behaviour at the terminal time* $t = T$. Evolution of such a conditional expectation is described by the Kolmogorov Backward Equation.

Theorem I.1.3 (Kolmogorov Backward Equation)

Let $u(t, x)$ be defined as

$$u(t, x) = \mathbb{E} [g(X(T)) | X(t) = x],$$

³i.e. Riemann integral has zero quadratic variation and so does any smooth function.

⁴this could be for example a default threshold.

⁵Alternating-direction implicit.

where $g(X(T)) = u(T, X(T))$ is a deterministic function of $X(T) \in \mathbb{R}^d$. Assuming that all necessary integrability and smoothness conditions are satisfied, u satisfies the Kolmogorov Backward Equation

$$\frac{\partial u(t, x)}{\partial t} + \mathcal{A}u(t, x) = 0, \quad (\text{I.1.4})$$

with the terminal condition

$$u(T, x) = g(x),$$

and \mathcal{A} being the infinitesimal generator

$$\mathcal{A} = \sum_{i=1}^d \mu_i(t, x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d [\sigma(t, x) \sigma(t, x)^T]_{i,j} \frac{\partial^2}{\partial x_i \partial x_j}. \quad (\text{I.1.5})$$

Since a contingent claim is a random variable with a certain behaviour (payoff) at its maturity T and the present value of this claim is a discounted expectation of the terminal outcome, the relevance of KBE with the option pricing is obvious.⁶ If we were to justify KBE mathematically, we would make use of the martingale property of the conditional expectation,⁷ and subsequently would apply Itô's lemma on u , as a function of t and X . This would yield a SDE for u whose drift by the martingale property (see also the Theorem I.1.1) must be equal to 0. The drift equated to 0 is then the KBE expression (I.1.4). In other words, KBE is a reflection of the zero drift condition of the martingale u .

An interesting problem arises when g is chosen to be

$$g(x) = \exp(i\omega^T x), \quad i = \sqrt{-1},$$

where ω is a dummy variable (or vector). Then

$$u(t, x) = \mathbb{E} [\exp(i\omega^T X(T)) | X(t) = x],$$

with the terminal condition $u(T, X(T)) = \exp(i\omega^T X(T))$. In such a setting, u is the (conditional) *characteristic function* of $X(T)$. Therefore, we postulate that even the characteristic functions of Itô's processes satisfy the KBE. Finally, KBE can also be used for the description of the evolution of probability density.⁸ Let us define a transition density function

$$p(t, x; s, y) dy = \mathbb{E} [X(s) \in [y, y + dy] | X(t) = x], \quad 0 \leq t \leq s \leq T. \quad (\text{I.1.6})$$

In the case of KBE, we analyze p evolving with respect to (t, x) given (s, y) is fixed. Precisely

$$\begin{aligned} \frac{\partial p(t, x)}{\partial t} + \mathcal{A}p(t, x) &= 0, \quad (s, y) \text{ fixed} \\ p(s, x; s, y) &= \delta(x - y), \end{aligned}$$

⁶notice that KBE does *not* include the discounting that appears in option pricing problems and hence we cautiously write that KBE is only 'related' to option pricing.

⁷recall $u(t, x) = \mathbb{E} [u(T, X(T)) | X(t) = x]$ and so by virtue of the Definition I.1.3 u is a martingale.

⁸this shall also be clear from its ability to describe the 'evolution' of a characteristic function. Later in the text it will be mathematically justified how the probability density function relates to a characteristic function.

where δ is the Dirac delta function. In view of the KBE, we might think of p being a density having a predefined behaviour at the future time s . This is captured by the terminal condition in terms of Dirac delta. If X is at the position y at time $t = s$, then the density is infinite at the singleton y , and 0 everywhere else. However, for some prior choice of $t < s$ the density function p will be 'wider'.

In a complementary fashion, we can define the Kolmogorov Forward Equation (KFE) as the logical opposite of KBE. Since KBE works back in time and produces 'historical' probability densities from known densities in the future, it is a little less practical for the use. Often, we are interested in how the 'current' density spreads with the future uncertainty in the system. This phenomenon is exactly described by the KFE.

Let \mathcal{A}^* denote the *adjoint* infinitesimal generator such that

$$\mathcal{A}^* f(s, y) = - \sum_{i=1}^d \frac{\partial (\mu_i(s, y) f(s, y))}{\partial y_i} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 \left(\left[\sigma(s, y) \sigma(s, y)^\top \right]_{i,j} f(s, y) \right)}{\partial y_i \partial y_j}.$$

Using \mathcal{A}^* we can define the Kolmogorov Forward Equation.

Theorem I.1.4 (Kolmogorov Forward Equation for transition density)

Assume p being a transition density as defined in (I.1.6) while (t, x) are fixed variables. Then p satisfies the Kolmogorov Forward Equation

$$\begin{aligned} - \frac{\partial p(s, y)}{\partial s} + \mathcal{A}^* p(s, y) &= 0, (t, x) \text{ fixed} \\ p(t, x; t; y) &= \delta(x - y). \end{aligned}$$

According to the boundary (in this case rather an initial) condition at $s = t$, $p(t, x; t; y) = \delta(x - y)$ it is obvious that the infinite density is concentrated at $y = x$ and as s moves towards the future time T , the density will spread. In other words, we know the system now (at $s = t$) and expect the KFE to explain its future behaviour in the sense of probability.

It is clear that both the KBE and the KFE in terms of the evolution of p describe the same phenomenon, each however from a different time direction.

Finally, to conclude this section, we can not forget to introduce the Feynman-Kac theorem which is frequently used to quickly obtain a PDE for the value of the derivative in terms of the state variables. It is a useful extension of the KBE. The main difference between KBE and Feynman-Kac is that Feynman-Kac already includes the 'discounting' which makes it application-ready for finance.

Theorem I.1.5 (Feynman-Kac)

Let $u(t, x)$ be defined as

$$u(t, x) = \mathbb{E} \left[\eta(t, T) g(X(T)) + \int_t^T \eta(t, s) h(s, X(s)) ds \middle| X(t) = x \right],$$

where

$$\eta(t, T) = \exp \left(- \int_t^T \zeta(s, X(s)) ds \right),$$

and x is $It\bar{o}$'s process as defined in (I.1.1). Then, modulo some integrability and regularity conditions, u solves

$$\frac{\partial u(t, x)}{\partial t} + \mathcal{A}u(t, x) + h(t, x) - \zeta(t, x)u(t, x) = 0,$$

with the terminal condition $u(T, x) = g(x)$ and infinitesimal generator \mathcal{A} given by (I.1.5).

The theorem is intentionally stated in terms of some general variables. It is not hard to see that η could be a discount factor and ζ is a placeholder for the instantaneous (and possibly stochastic) discounting rate. Indeed, it is easy to see that if $\zeta = r$, $h = 0$ and considering a process X with \mathbb{Q} -Wiener, the PDE turns out to be a PDE for option valuation under the spot martingale measure \mathbb{Q} . We note that the regularity issues are discussed in a great detail in (Duffie, 2010). For a PDE-related approach, Feynman-Kac result is an indispensable tool. It efficiently produces a valuation PDE given SDEs of the underlying processes.⁹ Feynman-Kac will play a critical role in the Section III.2 where particular pricing PDEs originate for claims subject to default intensity γ and also in the Section III.2.2.2 where we furthermore investigate the extension for jump diffusions and the related partial integro-differential equation (PIDE).

I.1.4 Self-financing trading strategies and arbitrage

Departing from the purely mathematical and probabilistic topics, in relation with contingent claims we need to investigate the concept of self-financing trading strategies and arbitrage. The key idea behind the valuation of derivatives is that the payoff can be replicated by a continuous trading in the primary assets. Therefore, it is useful to discuss these trading strategies.

Consider a market consisting of d assets X_1, \dots, X_d and assume that the process $X(t) = (X_1(t), \dots, X_d(t))^T$ is defined in (I.1.1). Also imagine a d -dimensional predictable¹⁰ stochastic process $h(t) = (h_1(t), \dots, h_d(t))^T$, where $h_i(t)$ represent holdings of i th asset at time t . We call h a *trading strategy*.

In terms of h we can construct a portfolio-value process π such that

$$\pi(t) = \sum_{i=1}^d h_i(t)X_i(t) = h(t)^T X(t),$$

and define a *self-financing* trading strategy.

Definition I.1.8 (Self-financing trading strategy)

A trading strategy h is self-financing if and only if

$$d\pi(t) = \sum_{i=1}^d h_i(t)dX_i(t) = h(t)^T dX(t). \quad (\text{I.1.7})$$

Self-financing portfolio means that no proceeds can be added or withdrawn from the

⁹the SDEs, however, need to represent the dynamics of the underlying under the martingale measure.

¹⁰without any rigour, a *predictable* process is a process whose value at time $(t + dt)$ is known at time t .

portfolio once one started with some initial value $\pi(0)$. We can also write (I.1.7) as

$$\pi(t) - \pi(0) = \int_0^t h(u)^\top dX(u) = \sum_{i=1}^d \int_0^t h_i(u) dX_i(u),$$

from which it is explicitly clear that the difference $\pi(t) - \pi(0)$ of the wealth of the portfolio only depends on the the trading gains and losses over time.

Clearly, we shall also examine under what conditions is the market arbitrage-free. For this the definition of the arbitrage is necessary.

Definition I.1.9 (Arbitrage)

An arbitrage by means of a self-financing trading strategy exists when

$$\pi(0) = 0, \pi(t) \geq 0, \mathbb{P}[\pi(t) > 0] > 0,$$

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

The first condition is the initial condition stating that the 'initial investment' is zero. The second condition implies that the wealth process π stays non-negative over $[0, T]$. The third condition ensures that there is a positive probability that the wealth π is positive. To conclude, an arbitrage is an operation with zero investment, in which the investor can not lose money and if he waits for a sufficiently long time, then this operation yields a profit.

I.1.5 Martingale measure and Girsanov theorem

A natural question is under which measure is the market arbitrage-free such that the introduced arbitrage (see the Definition I.1.9) is prevented. If we can find such a measure and subsequently show that the payoff of a derivative can be replicated by a self-financing portfolio, then we have a framework for the valuation of derivatives. The measure we are looking for is called the *equivalent martingale measure* (EMM) and it is the core component of the pricing of the contingent claims.

In order to derive the equivalent martingale measure, we firstly need to introduce the change of measure.

Theorem I.1.6 (Change of measure)

Fix a measurable space (Ω, \mathcal{F}) and let \mathbb{P} and \mathbb{Q} be two equivalent measures on (Ω, \mathcal{F}) . Let L be a nonnegative random variable such that $\mathbb{E}^\mathbb{P}[L] = 1$. For all $A \in \mathcal{F}$ define

$$\mathbb{Q}[A] = \mathbb{E}^\mathbb{P}[L\mathbf{1}_{\{A\}}].$$

Then \mathbb{Q} is a probability measure. Furthermore, if X is also a nonnegative random variable, then

$$\mathbb{E}^\mathbb{Q}[X] = \mathbb{E}^\mathbb{P}[LX].$$

Finally, if L is a strictly positive random variable and Y is a nonnegative random variable, then

$$\mathbb{E}^\mathbb{P}[Y] = \mathbb{E}^\mathbb{Q}\left[\frac{Y}{L}\right].$$

This theorem ensures that having a suitable random variable L , we can switch between measures \mathbb{P} and \mathbb{Q} . Focusing on the first part it is easy to see that probability of some event A under the 'new measure' \mathbb{Q} is equal to the probability of A under \mathbb{P} ,¹¹ distorted by some random variable L . The random variable L is called the *Radon-Nikodym* derivative and it is a tool that allows to switch between measures. The second point in the theorem is particularly useful for the financial applications. It expresses that expectation of some random variable under the new measure is equal to its expectation under the original measure, at the same time multiplied by the Radon-Nikodym derivative L .

In fact, the Radon-Nikodym derivative can be thought of as a density process which is captured by the following Theorem.

Theorem I.1.7 (Radon-Nikodym density process and derivative)

Let \mathbb{P} and \mathbb{Q} be two equivalent measures on the measurable space (Ω, \mathcal{F}) and let L be a nonnegative random variable. Then the process

$$L(t) = \mathbb{E}^{\mathbb{P}} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \middle| \mathcal{F}(t) \right] = \mathbb{E}^{\mathbb{P}} [L(T) | \mathcal{F}(t)], t \in [0, T],$$

is a \mathbb{P} -martingale and is called the *Radon-Nikodym density (or derivative) process*, with the property $L(0) = 1$. For a fixed T , the quantity

$$L(T) = \frac{d\mathbb{Q}}{d\mathbb{P}},$$

is referred to as the *Radon-Nikodym derivative*.

It is obvious that Theorem I.1.6 and Theorem I.1.7 form a set to change the expectations. The main result is the fact that having the Radon-Nikodym derivative L gives a tool to easily switch the measures (and thus also expectations under these measures).

Having outlined the change of measure it is also useful to introduce EMM.

Theorem I.1.8 (Equivalent Martingale Measure)

Let Y be a deflator (asset) and consider the deflated asset process $\tilde{X}(t) = (X_1(t)/Y(t), X_2(t)/Y(t), \dots, X_d(t)/Y(t))$. The measure that renders the deflated price process \tilde{X} into a martingale is the *equivalent martingale measure (EMM)*.

Notice that this theorem can also be reversed and so under the measure associated with the deflator asset Y , \tilde{X} is *guaranteed* to be a martingale. Also, if the asset Y is one of the assets in X , e.g. X_1 , then X_1 is referred to as the *numéraire* asset.

Going back to the self-financing trading strategies, we say the trading strategy h is *permissible*,¹² if and only if

$$\int_0^t h(u)^T d\tilde{X}(u),$$

is a martingale under the EMM associated with the deflator asset (or possibly numéraire).

¹¹note that $\mathbb{P}[A] \triangleq \mathbb{E}^{\mathbb{P}}[\mathbf{1}_{\{A\}}]$.

¹²some authors also use the word *admissible*. Admissible/permissible strategies, put crudely, are strategies whose generated wealth π is a.s. bounded from below.

This in particular means in the world of Itô's processes that in order to make h a permissible strategy, the process¹³ $h(t)^\top \sigma_{\tilde{X}}(t)$ must be in H^2 (see Definition I.1.2). This condition ensures that the integral above is a martingale (not only a local martingale) and so some 'wild' trading strategies are ruled out.¹⁴

The concept of pricing under EMM requires that all deflated processes are martingales under the EMM. The only point that still has not been explained is *how* to change the measure so that the process becomes a martingale under the EMM. This specific point is addressed by the Girsanov theorem which provides a concrete formula for the measure change in the Itô's world.¹⁵

In what follows, we denote \mathbb{P} the original measure and \mathbb{Q} the new measure and the objective is to transform a process X under \mathbb{P} into a process X under \mathbb{Q} . To find X under \mathbb{Q} we firstly establish Radon-Nikodym density process L such that

$$dL(t) = L(t)\varphi(t)^\top dW^\mathbb{P}(t), \quad (\text{I.1.8})$$

where φ is the p -dimensional *transformation kernel* and $W^\mathbb{P}$ is a standard Wiener process under \mathbb{P} . This is the explicit representation of $L(t) = \mathbb{E}^\mathbb{P} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \middle| \mathcal{F}(t) \right]$. Solution for L is obviously an exponential martingale and it yields¹⁶

$$\begin{aligned} L(t) &= \exp \left(\int_0^t \varphi(u)^\top dW^\mathbb{P}(u) - \frac{1}{2} \int_0^t \varphi(u)^\top \varphi(u) du \right) \\ &= \varepsilon(\varphi * W^\mathbb{P})(t), \end{aligned}$$

where ε is the *Doléans exponential*. For L to be a true martingale (and not only a local martingale) it is then required that

$$\mathbb{E}^\mathbb{P} \left[\frac{1}{2} \int_0^t \varphi(u)^\top \varphi(u) du \right] < \infty.$$

This sufficient true martingale condition is named Novikov condition.¹⁷ After going through the technical issues, we can finally postulate an important result.

Theorem I.1.9 (Girsanov theorem)

Assume the existence of the Radon-Nikodym density process L equipped with a p -dimensional transformation kernel φ and furthermore suppose the Novikov condition holds. Then

$$dW^\mathbb{P}(t) = \varphi(t)dt + dW^\mathbb{Q}(t),$$

where $W^\mathbb{Q}$ is a p -dimensional standard Wiener process under \mathbb{Q} .

¹³by $\sigma_{\tilde{X}}$ we instantly denote the term $\sigma(t, X(t))$ of the deflated process \tilde{X} .

¹⁴one such a good example of a wild trading strategy might be the *doubling strategy* pointed out by (Harrison and Kreps, 1979).

¹⁵by Itô's world we mean a world, where all processes are Itô's.

¹⁶note that $\int_0^t \varphi(u)^\top dW^\mathbb{P}(u) = \int_0^t \sum_{i=1}^p \varphi_i(u) dW_i^\mathbb{P}(u) = \sum_{i=1}^p \int_0^t \varphi_i(u) dW_i^\mathbb{P}(u)$.

¹⁷validity of this condition might be very difficult to verify in practice and so usually it is implicitly assumed that the Novikov condition is in effect.

The Girsanov theorem is a powerful theorem which allows to switch the dynamics of the Itô's processes with respect to the measure in consideration. The change of the measure is ensured by the Girsanov transformation kernel φ .

I.1.6 Pricing of contingent claims and the change of numéraire

The main objective is the pricing of contingent claims. Having outlined the technical details we shall come up with general formulas for pricing of contingent claims. It is natural to start with two fundamental theorems of asset pricing. We present simplified, yet still general versions. More exhaustive formulations (with proofs) can be found in (Björk, 2009, Section 3), (Andersen and Piterbarg, 2010, pp. 10-11) or almost encyclopedically in (Musiela and Rutkowski, 2010).

Theorem I.1.10 (No arbitrage - First Fundamental Theorem)

If there exists at least one EMM such that deflated processes follow martingales, then there is no arbitrage.

Theorem I.1.11 (Completeness - Second Fundamental Theorem)

If the EMM is unique, then the market is complete.

The first theorem is effectively a *sufficient* condition for no-arbitrage. However, no-arbitrage is not the same thing as the existence of the EMM. Despite this 'small' difference we treat the existence of EMM and no-arbitrage as two similar concepts.¹⁸ In contrast, no-arbitrage does not imply completeness. The second theorem states that in a complete market, prices of derivatives must be unique (because of the uniqueness of the EMM). Notice that the completeness can also be equivalently defined via replicability of all derivatives. If all derivatives can be replicated, then the market is complete. It is common in the pricing of derivatives to assume that the second theorem is in effect although in practice it might not be true. An excellent classification of models (as arbitrage-free and/or complete) is provided by (Björk, 2009, Meta-theorem 8.3.1 and Section 15).

We shall now glue these two theorems together and provide a formula for valuation. To this end, it is convenient to start with some derivative-construction arguments and based on these arguments derive the so-called *risk-neutral valuation formula*. Assume there is a fixed time horizon T and that there exists a derivative security paying out $V(T)$ at time T . If there exist a permissible trading strategy h for which $V(T) = \pi(T)$, then this derivative is *attainable*.¹⁹ We say that the trading strategy h *replicates* the derivative. By virtue of the arbitrage argument (see Definition I.1.9), the initial cost of such a replicating portfolio *must* be equal to the initial value of the derivative security $V(0)$. Therefore, $V(0) = \pi(0)$ must necessarily hold and so does $V(t) = \pi(t), t \in [0, T]$.²⁰ Let N be a numéraire (or deflator) asset and let \mathbb{Q}_N the related martingale measure. Employing the First Fundamental Theorem leads to the observation that elimination of the arbitrage

¹⁸see (Björk, 2009, section 10.2.1) for the proof of how the existence of EMM implies no-arbitrage and vice versa.

¹⁹in some literature, this is also referred to as *redundant*.

²⁰if the equality did not hold, yet still the trading strategy h with initial set up cost π would replicate $V(T)$, then it is clear an arbitrage could be achieved.

requires a martingale measure under which the deflated assets are martingales which in the case of the replicating portfolio and the derivative value means

$$\frac{\pi(t)}{N(t)} = \mathbb{E}^{\mathbb{Q}_N} \left[\frac{\pi(T)}{N(T)} \middle| \mathcal{F}(t) \right] = \mathbb{E}^{\mathbb{Q}_N} \left[\frac{V(T)}{N(T)} \middle| \mathcal{F}(t) \right] = \frac{V(t)}{N(t)}.$$

Taking out just the V -related equality and rearranging yields the *risk-neutral valuation formula*

$$V(t) = N(t) \mathbb{E}^{\mathbb{Q}_N} \left[\frac{V(T)}{N(T)} \middle| \mathcal{F}(t) \right]. \quad (\text{I.1.9})$$

In some simple pricing problems, the equation (I.1.9) is directly solvable. In many other situations, it might be not, although the dynamics under \mathbb{Q}_N has already been established (but the expectation does not have a simple solution). If one is about to derive an analytical formula (rather than a PDE), then he can possibly change the numéraire to obtain a simpler formula. This approach can in some specific cases turn ugly-looking expression into simple ones.

The change of numéraire can be derived as follows. Assume that the EMM \mathbb{Q}_N associated with the numéraire asset N is known and one would like to change the numéraire to some asset M from N . We firstly observe that at time $t = 0$

$$\frac{V(0)}{N(0)} = \mathbb{E}^{\mathbb{Q}_N} \left[\frac{V(T)}{N(T)} \right],$$

but because the risk-neutral valuation formula is valid for any numéraire, it must also hold

$$\frac{V(0)}{M(0)} = \mathbb{E}^{\mathbb{Q}_M} \left[\frac{V(T)}{M(T)} \right].$$

If we in each formula find only expression for V and equate these, then

$$\mathbb{E}^{\mathbb{Q}_M} \left[\frac{M(0)V(T)}{M(T)} \right] = \mathbb{E}^{\mathbb{Q}_N} \left[\frac{N(0)V(T)}{N(T)} \right].$$

The main trick comes from the second point of the Theorem I.1.6. According to the point, with $\mathbb{Q} = \mathbb{Q}_M$ and $\mathbb{P} = \mathbb{Q}_N$, we also obtain

$$\mathbb{E}^{\mathbb{Q}_M} \left[\frac{M(0)V(T)}{M(T)} \right] = \mathbb{E}^{\mathbb{Q}_N} \left[\frac{N(0)V(T)}{N(T)} \right] = \mathbb{E}^{\mathbb{Q}_N} \left[L(T) \frac{M(0)V(T)}{M(T)} \right],$$

where we additionally have the extra Radon-Nikodym derivative L . If we now compare the terms under the $\mathbb{E}^{\mathbb{Q}_N}[\cdot]$ operator, we find

$$L(T) = \frac{M(T)N(0)}{M(0)N(T)} \Rightarrow L(t) = \frac{M(t)N(0)}{M(0)N(t)}.$$

This is the Radon-Nikodym derivative L , whose differential reveals the explicit form of the Girsanov kernel φ . This kernel can then be used in order to switch $W^{\mathbb{Q}_N}$ to $W^{\mathbb{Q}_M}$ as stated by the Theorem I.1.9. We summarize the change of numéraire by a theorem.

Theorem I.1.12 (Change of numéraire)

Let $\mathbb{Q}_N, \mathbb{Q}_M$ be two equivalent martingale measures corresponding to some numéraire assets N and M , respectively. Assume that the dynamics of the market is well-known

under \mathbb{Q}_N and one would like to price under \mathbb{Q}_M . Then, the Radon-Nikodym derivative for switching from \mathbb{Q}_N to \mathbb{Q}_M is equal to

$$L(t) = \frac{M(t)N(0)}{M(0)N(t)}.$$

Corollary I.1.3 (Girsanov kernel in numéraire change for geometric processes)
Given the setup and the result of the Theorem I.1.12, under the assumption that the processes under \mathbb{Q}_N follow

$$\begin{aligned} dN(t) &= N(t)\mu_N(t)dt + N(t)\sigma_N(t)dW^{\mathbb{Q}_N}(t) \\ dM(t) &= M(t)\mu_M(t)dt + M(t)\sigma_M(t)dW^{\mathbb{Q}_N}(t), \end{aligned}$$

where σ_N, σ_M are row vectors, the Girsanov transformation kernel equals

$$\varphi(t) = \sigma_M(t) - \sigma_N(t).$$

This theorem provides a very useful shortcut which can be directly combined with the Girsanov transformation (see Theorem I.1.9) and thus one quickly obtains the dynamics under the 'new' martingale measure.

Proof. The proof resides in the calculation of the differential of L and extracting φ by comparison with (I.1.8). Observe that

$$dL(t) = \frac{N(0)}{M(0)} d\left(\frac{M(t)}{N(t)}\right) = \frac{N(0)}{M(0)} d\left(M(t)\frac{1}{N(t)}\right).$$

Then find

$$d\left(\frac{1}{N(t)}\right) = -\frac{1}{N(t)}\left(\mu_N(t)dt + \sigma_N(t)dW^{\mathbb{Q}_N}(t) - \sigma_N(t)\sigma_N(t)^T dt\right).$$

Use the Itô's product rule (I.1.3) to compute $d\left(M(t)\frac{1}{N(t)}\right)$ and so

$$d\left(M(t)\frac{1}{N(t)}\right) = \frac{1}{N(t)}dM(t) + M(t)d\left(\frac{1}{N(t)}\right) + dM(t)d\left(\frac{1}{N(t)}\right).$$

Substituting for $dN(t)$, $dM(t)$ and $d\left(M(t)\frac{1}{N(t)}\right)$ gives a relatively complicated expression which can be compactly written as

$$d\left(\frac{M(t)}{N(t)}\right) = \frac{M(t)}{N(t)}(\dots)dt + \frac{M(t)}{N(t)}(\sigma_M(t) - \sigma_N(t))dW^{\mathbb{Q}_N}(t).$$

Yet because L is a \mathbb{Q}_N -martingale (see Section I.1.5) and every martingale by means of the Martingale Representation Theorem (see Theorem I.1.1) drops the dt -term, then simply

$$d\left(\frac{M(t)}{N(t)}\right) = \frac{M(t)}{N(t)}(\sigma_M(t) - \sigma_N(t))dW^{\mathbb{Q}_N}(t).$$

Using this result in the first formula for dL yields

$$dL(t) = \frac{N(0)}{M(0)} d\left(\frac{M(t)}{N(t)}\right) = \frac{N(0)}{M(0)} \frac{M(t)}{N(t)} (\sigma_M(t) - \sigma_N(t)) dW^{\mathbb{Q}_N}(t).$$

However, since $\frac{M(t)N(0)}{M(0)N(t)} = L(t)$, then this is

$$dL(t) = L(t) (\sigma_M(t) - \sigma_N(t)) dW^{\mathbb{Q}_N}(t).$$

Comparing with (I.1.8) it is easy to see that the Girsanov transformation kernel equals

$$\varphi(t) = \sigma_M(t) - \sigma_N(t).$$

□

I.1.7 Practical observations and comments

In the previous sections we have provided a set of tools for handling the stochastic processes in the Wiener world and also tools for the pricing of contingent claims. With all these tools one can readily price contingent claims. However, the focus was kept on the general statements, and all practical comments were kept until this section. In this part we therefore provide several useful subjective remarks related to the applications rather than to some objective explanation of the subject.

First of all, we shall comment the 'usual' conventions in the naming of measures. Typically, one assumes there is a physical (real-world or statistical) measure, which we henceforth label \mathbb{P} , although the measure \mathbb{P} in the former section denoted some general, unspecified measure. The risk-neutral measure, in our case, means a general EMM. Notice that some authors directly associate the term 'risk-neutral measure' with a specific type of EMM, commonly the spot martingale measure. We therefore define the *spot martingale measure* as a measure that makes all money market B -discounted assets martingales. We denote this measure henceforth \mathbb{Q} . In many practical applications, particularly in the fixed-income area, it is convenient to work under the T -forward measure \mathbb{Q}_T induced by the numéraire T -bond $B(\cdot, T)$.²¹

Our next remark is related to the valuation itself if only the dynamics under \mathbb{P} is known. The typical approach is the following. Find a candidate numéraire asset (see below what assets are typically chosen as a numéraires). Then find the EMM that makes all numéraire-deflated assets martingales under the EMM. This can be done by application of Itô's lemma (Theorem I.1.2) on the deflated process. Identify the Girsanov kernel φ that makes the deflated-processes martingales (under the EMM) and apply the Girsanov theorem (Theorem I.1.9) to change the measure of all processes from \mathbb{P} to the EMM associated with the numéraire. Then the valuation formula (I.1.9) can be established for that specific valuation problem.

Our third comment is to the application of the numéraire change which is formally presented by the Theorem I.1.12. There are several typical situations *when* one might want to change the numéraire. The first situation is when the dynamics is only specified under some other EMM (connected with some prespecified numéraire). Commonly, dynamics of various processes is specified under the spot martingale measure \mathbb{Q} but this might be an unsuitable choice for many applications. Secondly, after choosing a 'good' numéraire,

²¹the reason is purely practical. Typically, working under \mathbb{Q}_T eliminates a stochastic discount factor in the risk-neutral valuation formula (I.1.9), which would normally be present in the valuation under \mathbb{Q} when the interest rates are stochastic.

when the risk-neutral valuation formula (I.1.9) is simplified, one might encounter 'new' expectations which might not be trivial to solve. In such a scenario, one is free to conduct a change of numéraire just for the expectation which would be difficult to solve under the prior EMM associated with the 'good' numéraire. This case, for example appears in Black-Scholes when the 'good' numéraire is, say, B , but then one needs to change the measure to a 'stock measure', where the *stock* under \mathbb{Q} is the numéraire. See (Musielà and Rutkowski, 2010, p. 130) for the illustration.

Cautious decisions need to be made upon the *choice* of numéraire. It was already discussed how to switch the numéraire yet still, it was not discussed what are possible choices of the numéraire. In the most applications, the money market account B or the T -bond $B(\cdot, T)$ are selected as numéraires. This is, however, suboptimal for many pricing problems. For example, in pricing of swaptions, it is common to use annuity²² as a numéraire (see e.g. (Kwok, 2008) or (Hippler, 2008)²³). Another good example might be the exchange option (Margrabe, 1978), where one of the stocks is chosen as the numéraire and hence the money-market account does not play any role at all. More importantly in the FX market one often chooses the foreign money market account as a numéraire (see (Garman and Kohlhagen, 1983)). Obviously, the choice of the numéraire is not arbitrary and the reader is referred to an excellent text (Björk, 2009, Section 26), which provides exhaustive explanation of the choice (and the change of) numéraires. In this text, we nearly in all the cases work with processes under the risk-neutral measure \mathbb{Q} . Only in several, exceptional cases, other choices of measure are used.

Our final words defend the presence of the risk-free²⁴ rate in the financial models. In the most applications some risk-free rate appears although arguably the markets are not risk-free as counterparties can default on their liabilities (possibly also emerging from contingent claims). This might lead to the conclusion that risk-free rate shall be replaced by a different discounting rate. However, after the 2007-2008 crisis it became a standard between counterparties to collateralize their transactions (by means of the Credit Support Annex) and this collateralization under suitable set of conditions makes the deals *virtually* default free. Risk-free rate therefore is in place. For uncollateralized deals the risk-free rate also defends its relevance. This is because the value of such a deal is actually calculated as a value of risk-free deal, adjusted for CVA and DVA (+ possibly FVA).²⁵ In calculation of the value of the risk-free deal, risk-free rate needs to be used.

²²annuity here means a portfolio of discount bonds.

²³the author does not explicitly name the measure 'annuity measure', he only refers to a portfolio of discount bonds as numéraire. These are, however, equivalent terms.

²⁴risk-free has the meaning of default-free.

²⁵CVA = credit valuation adjustment, DVA = debit valuation adjustment, FVA = funding valuation adjustment.

Part II

Firm–value models

Having origins in the 1970s with the rise of the Black-Scholes formula, the firm-value models represent a classical way of credit risk modeling, in that they try to model the default as a process of declining value of the firm’s assets which is usually the case of default in the real-world. Because later (Part III) we investigate in a great detail the reduced-form models, we find it highly informative to include the firm-value part too such that a comparison is possible between the two distinct approaches. Our objective in this Part is to provide a consistent exposition on the firm-value models, derive important pricing PDE and conduct an analysis of the *pricing* performance of the firm-value models.

This Part II is organized as follows. In the Sections II.1.1–II.1.3 we describe the general modeling environment whose main outcomes are the general risk-neutral valuation formula and the associated pricing PDE which we completely derive. Section II.1.4 and the related subsections provide discussion of some established models. In the Section II.1.5 we conduct an analysis which aims to prove (or reject) the hypothesis whether a firm-value model calibrated using the stock market data can be outperformed by a model that uses option market data. The conclusion is that the option market provides data for a more sensitive credit analysis than the stock market.

II.1 Firm–value approach to credit risk

The firm-value models in the credit risk modeling are generally regarded as the older class of credit risk models, compared to reduced-form models discussed in a great detail in the Part III. Firm-value models were pioneered by Black, Scholes and Merton with the discovery of the Black-Scholes model. The main idea behind the firm-value approach is self-explanatory. These models capture the credit risk by means of changes in the structure of the balance sheet of a firm. This is fundamentally underpinned by the empirical observations of the most defaults being caused by insufficient assets to meet the liabilities.

In this part of the text, we provide a complete specification of the general firm-value modeling environment in which the main component is the general (risk-neutral) valuation formula for defaultable claims. This subsequently leads to a general pricing PDE for defaultable claims under the firm-value approach which we derive by combining the expectation operators defining the price of the components. After these two general sections, we discuss several popular structural models, in particular the classical model of (Merton, 1974) (based on (Black and Scholes, 1973)), the jump diffusion model of (Zhou, 1997) (based on (Merton, 1976)), and the stochastic volatility model (Heston, 1993). We also derive a closed-form formula for Black-Scholes with stochastic rates which possibly better suits for pricing of interest-rate sensitive contingent claims. The empirical part is then dedicated to the analysis of the prediction power of these models. We, in particular, calibrate these models using the quoted option prices, that is, we also investigate whether option quotes carry better *credit* information than just the stock market. We find that the option market provides more accurate credit information than the stock market.

II.1.1 Preliminaries and the general modeling environment

Because most of the firm-value models are based on a rather similar principle, it is beneficial to define some general modeling environment which we specialize for each particular model. With this idea in mind, it will be shown how the risk-neutral valuation formula (I.1.9) adjusts according to model specification. We therefore do not present the firm-value models separately (and independently) but follow quite a uniform approach such that there is a better linkage.

We assume the existence of the following variables and processes:

- asset value process X . This process will represent a *hypothetical* process of the firm value,
- cumulative *promised* dividend process A which captures all promised dividend-like cash-flow (discrete and/or continuous) streaming from the defaultable claim, passed to the claim holder,
- recovery process Z . At default, the holder of the claim receives Z ,
- recovery claim δ . If default occurs prior the claim's maturity T , the claimholder receives δ at T ,
- cumulative *realized* cash-flow process C . This process is subject to the promised dividends A , recovery process Z and the recovery claim δ ,
- contingent T -terminal payment g , possibly being dependent on the assets X and/or the risk-free rate r ,
- risk-free rate process r ,
- default-barrier process h ,
- money-market account B whose value is fully determined just by the risk-free rate r ,
- default-free zero bond $B(\cdot, T)$ maturing at time T ,
- time of default τ^* .

We also consider a probability space $(\Omega, \mathcal{G}, \mathbb{P})$ with a filtration $\mathbb{G} = \{\mathcal{G}(t)\}_{0 \leq t \leq T^*}$ which is capable to capture all these processes and variables. In order to formally define the default time, allow us firstly the following definition.

Definition II.1.1 (\mathbb{G} -stopping time)

Assume the filtration \mathbb{G} . Then the random time τ^* is \mathbb{G} -stopping time if

$$\{\tau^* \leq t\} \in \mathcal{G}(t).$$

We will for short refer to any τ^* satisfying this definition as \mathbb{G} -stopping time. Heuristically, the definition postulates that at time t we are able to determine if the event $\{\tau^* \leq t\}$ occurred or not, given the random variable τ^* depends on a value of some random process that generated or is adapted to the filtration \mathbb{G} . Advancing further, it is also useful to introduce a *predictable* stopping time.

Definition II.1.2 (Predictable stopping time)

If there exists an increasing sequence of stopping times $\tau_1^*, \tau_2^*, \dots, \tau_n^*$ such that

$$\lim_{n \rightarrow \infty} \tau_n^* = \tau^* \text{ a.s.},$$

then the stopping time τ^* is predictable. The increasing sequence $\tau_1^*, \tau_2^*, \dots, \tau_n^*$ is said to announce the stopping time τ^* .

Having defined the predictability of a stopping time, we can finally also specify the time of default in a typical firm-value model.

Definition II.1.3 (Time of default in a firm-value model)

Given a default barrier h the time of default τ^* in a firm-value model is a \mathbb{G} -stopping time such that

$$\tau^* = \inf \{t \in \mathcal{T} : X(t) \leq h(t)\},$$

where \mathcal{T} is a set of all permissible times when default can occur. Furthermore, if the filtration \mathbb{G} is generated by a (possibly multidimensional) Wiener process (only), then τ^* is \mathbb{G} -predictable stopping time. If X involves pure jumps then τ^* loses its predictability property.

It is characteristic for most of the firm-value models that τ^* is \mathbb{G} -predictable. On the intuitive level, processes driven exclusively by the Wiener process generate predictable stopping times since the Wiener process is a continuous process. Therefore, the default can *not* come as a surprise, since there exists a sequence of increasing stopping times that announce the default time τ^* . Let us give an example of the announcing sequence. Consider $X(0) > h$, where h is for simplicity some fixed default barrier. Then the announcing sequence of stopping times can be

$$\tau_i^* = \inf \left\{ t \in \mathbb{R}^+ : X(t) \leq h + \frac{1}{i} \right\}, i = 1, 2, \dots, n.$$

Taking the limit $n \rightarrow \infty$ implies $\tau_n^* \rightarrow \tau^*$ a.s., as required by the Definition II.1.2. If, however, the process X involves jumps, then τ^* will still remain \mathbb{G} -stopping time, yet it will *no longer* be \mathbb{G} -predictable stopping time. A rigorous discussion of stopping times is provided by (Protter, 2005, Section III (2)).

A natural question is whether there is or is not a conflict between the recovery claim δ and the recovery process Z . Apparently, both elements can not be present at the same time in the valuation for the given contingent claim because there is always only one recovery policy in place. Therefore, either δ or Z is in force but not both simultaneously. Inclusion of both δ and Z in the list of elements above serves solely for a complete specification of the modeling environment such that all possibilities are covered. To support this point with explicit examples, consider a defaultable claim under the Fractional Recovery of Treasury Value (FRTV). In such a case, a recovery fraction δ is received at bond's maturity T , regardless of the default time $\tau^* \leq T$ (see e.g. (Jarrow and Turnbull, 1995)²⁶ or (Bingham and Kiesel, 2013, p. 395)). This would involve employing δ and assuming away the process Z . In the general formula that follows, we, however, retain both δ and Z in the formula.

²⁶this paper pioneered the FRTV recovery rule.

The 'unused' policy component in such a formula is then set to 0 to effectively disregard such a policy.

II.1.2 General risk-neutral valuation formula

In the Section I.1.6 we outlined the risk-neutral valuation formula (I.1.9). This formula, however, was only suitable for contingent claims with a single payoff at time T . When pricing contingent claims involving credit risk, the formula (I.1.9) will prove to be insufficient. The main reason for this deficiency is that the defaultable claims typically involve some intermediate cash-flow,²⁷ not only a single payoff at maturity T and also that there could be some recovery settlement at the default time τ^* . Therefore, we need an extended version of the former risk-neutral valuation formula. However, to build up such a formula, it is first necessary to provide a more concrete specification of the *realized* cash-flow process C which is an absolutely key component. Notice that we do not assume any possible implicit credit migration as the component of the credit risk here, credit risk is only modeled in terms of a default event.²⁸

For the valuations, we introduce the following essential processes: the default process H and the realized cash-flow process C . Combining these two processes, we can finally establish the enhanced risk-neutral valuation formula for the defaultable claims. Two definitions follow.

Definition II.1.4 (Default process)

The default process H corresponds to the process of default indicator

$$H(t) = \mathbf{1}_{\{\tau^* \leq t\}}.$$

Definition II.1.5 (Cash-flow process)

The generalized cumulative cash-flow process C of a defaultable claim equals

$$C(t) = (\delta \mathbf{1}_{\{\tau^* \leq T\}} + g \mathbf{1}_{\{\tau^* > T\}}) \mathbf{1}_{\{t \geq T\}} + \int_0^t (1 - H(u)) dA(u) + \int_0^t Z(u) dH(u).$$

We also make the following 'continuous compounding' assumption within the promised dividend process A .

Proposition II.1.1 (Promised dividend process)

The process A of the promised cumulated dividends accrues at the continuous rate c such that

$$A(t) = \int_0^t c(u, r, X) du,$$

²⁷these are *not* increments of the process A . The process A represents a process of cumulative *promised* dividends A . The actual, realized cash-flow (and its differences) are modeled via the process C (see below the Definition II.1.5).

²⁸sometimes the critics of the firm-value models stress out this argument as a severe imperfection. In fact, it is possible to establish a methodology to map components of the model to some real-world probabilities (and thus also to credit ratings). A commercially successful product KMV (Crosbie and Bohn, 2003) has a large credit database which allows to map KMV's firm-value model to real-world default probabilities, so called *expected default frequencies* (EDFs).

where r is a risk-free rate and X is a firm-value process.

We stated this Proposition to highlight the fact that although c is a continuously compounded rate it can itself still be a discontinuous process. Finally, having defined H, C and having made the continuous compounding assumption, we can focus on the risk-neutral valuation formula for defaultable claims. This the objective of the two definitions that follow.

Definition II.1.6 (Ex-dividend value)

The ex-dividend value $V(t, T)$ of a defaultable T -claim V is a process taking into account only the cash-flows paid by the claim in the future time period $(t, T]$.

Definition II.1.7 (Risk-neutral valuation formula for defaultable claims)

The ex-dividend value of a T -claim V which provides its claimholder with the cash-flow stream C is equal to

$$V(t, T) = B(t)\mathbb{E}^{\mathbb{Q}} \left[\int_t^T \frac{1}{B(u)} dC(u) \middle| \mathcal{G}(t) \right] = \mathbb{E}^{\mathbb{Q}} \left[\int_t^T \exp \left(- \int_t^u r(s) ds \right) dC(u) \middle| \mathcal{G}(t) \right], \quad (\text{II.1.1})$$

for $t \in (0, T)$. At maturity T then necessarily

$$V(T, T) = \delta \mathbf{1}_{\{\tau^* \leq T\}} + g \mathbf{1}_{\{\tau^* > T\}}.$$

Let us allow a few comments about the above stated definitions. The process of the default indicator H serves as a simple, yet practical tool when one is only concerned about the default and survival states of the credit. The realized cash-flow process C is combined from several components we shall investigate further. We stress out that C is assumed to be on a *cumulative* basis. Thus, $C(t) - C(u), u < t$ represents the *realized* total cash-flow²⁹ from the claim in the time interval $(u, t]$. The indicator $\mathbf{1}_{\{t \geq T\}}$ in C ensures that the round brackets terms are only in force at (and after) claim's maturity T .³⁰ If the default has occurred before the claim's maturity T , then it will pay off δ at T , otherwise it pays g . Let us now further analyze the two integrals in C . The first integral captures what dividend cash-flow streaming from the claim has been actually accumulated until the time t . To see this, observe that

$$\int_0^t (1 - H(u)) dA(u) = \int_0^t \mathbf{1}_{\{\tau^* > u\}} dA(u) = A(\tau^* -) \mathbf{1}_{\{\tau^* \leq t\}} + A(t) \mathbf{1}_{\{\tau^* > t\}}.$$

This is because

$$\int_0^t \mathbf{1}_{\{\tau^* > u\}} dA(u) = \begin{cases} \int_0^t dA(u) = A(t) & \text{if } \tau^* > t \\ \int_0^{\tau^* -} dA(u) = A(\tau^* -) & \text{if } \tau^* \leq t \end{cases}.$$

²⁹the cash-flow stream can also be negative. An example of the negative cash-flow might be a protection buyer CDS contract who *pays* a coupon (premium) for a protection.

³⁰this assumption of 'after maturity' is not a breach of any rules. Since C is on a cumulative basis, the process C does not grow further after T is reached.

This integral therefore captures the phenomenon that the dividends are paid until τ^*- and for any later time there is no further stream. The promised dividend process is thus *stopped*³¹ at $\min[\tau^*, t]$.

Finally the integral

$$\int_0^t Z(u)dH(u) = Z(\tau^*)\mathbf{1}_{\{\tau^* \leq t\}},$$

represents the immediate settlement payoff at default. Again, the last expression can be derived by fixing some τ^* and analyzing the integral on the two disjoint sets $\{\tau^* > t\}$ and $\{\tau^* \leq t\}$. It results into

$$\int_0^t Z(u)dH(u) = \begin{cases} \int_0^t Z(u)d\mathbf{1}_{\{\tau^* \leq u\}} = 0 & \text{if } \tau^* > t \\ \int_0^t Z(u)d\mathbf{1}_{\{\tau^* \leq u\}} = \int_0^{\tau^*} Z(u)d\mathbf{1}_{\{\tau^* \leq u\}} = Z(\tau^*) & \text{if } \tau^* \leq t \end{cases}.$$

Here, bear in mind that the integral over the indicator is automatically 0 as long as the value of the indicator remains constant on the integration interval. Here, in general notice that an integral in which the default time indicator H represents the *integrator*, over some general integrable function f , can be written as

$$\int_0^t f(u)d\mathbf{1}_{\{\tau^* \leq u\}} = f(\tau^*)\mathbf{1}_{\{\tau^* \leq t\}},$$

because the value of the integral in fact is equal to the value of f evaluated at the singleton time at which the indicator *changes* from 0 to 1 which can only occur at the time τ^* . It thus turns out that the expression for C in the Definition II.1.5 can be reduced to the integral-free expression

$$C(t) = (\delta\mathbf{1}_{\{\tau^* \leq T\}} + g\mathbf{1}_{\{\tau^* > T\}})\mathbf{1}_{\{t \geq T\}} + A(\tau^*)\mathbf{1}_{\{\tau^* \leq t\}} + A(t)\mathbf{1}_{\{\tau^* > t\}} + Z(\tau^*)\mathbf{1}_{\{\tau^* \leq t\}}.$$

After exploring the components of C the risk-neutral valuation formula (II.1.1) has an obvious interpretation. The value of a (not necessarily³²) defaultable T -claim V equals the \mathbb{Q} -expected money market B -discounted stream of the expected future cash-flow.

One might also wonder why the promised T -maturity payoff $(\delta\mathbf{1}_{\{\tau^* \leq T\}} + g\mathbf{1}_{\{\tau^* > T\}})$ is not included in the promised dividend process A . The main reason for this separate treatment is that typically, dividends (coupons) are subject to a different recovery policy than the notional. Often when a bond defaults, the bondholders only receive a fraction of the bond's face value and the interest (coupons) is assumed away. Therefore exclusion of the T -payoff from A allows for a more careful treatment of the recovery policies.

Finally, we note that the risk-neutral valuation formula (II.1.1) can be written in terms of the individual components.

³¹notice that any process X is a stopped process if and only if $X(t) = X(\min[\tau^*, t])$, where τ^* is a stopping time. This means that at the stopping time τ^* the process X freezes at the value $X(\tau^*)$ and remains at this value for any later time $t \geq \tau^*$.

³²as the formula (II.1.1) is very general, we might e.g. set $\delta = 1$, $A = 0$, $Z = 0$, $g = 1$ and the formula produces an expression for pricing of a risk-free zero bond $B(t, T)$.

Corollary II.1.1 (Risk-neutral valuation formula for defaultable claims)

The risk-neutral valuation formula (II.1.1) in terms of g , A , δ , and Z equals

$$\begin{aligned}
 V(t, T) &= \mathbf{1}_{\{\tau^* \leq t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T r(u) du \right) \delta \middle| \mathcal{G}(t) \right] \\
 &+ \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T r(u) du \right) \delta \mathbf{1}_{\{\tau^* \leq T\}} \middle| \mathcal{G}(t) \right] \\
 &+ \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T r(u) du \right) g \mathbf{1}_{\{\tau^* > T\}} \middle| \mathcal{G}(t) \right] \\
 &+ \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{\tau^*} r(u) du \right) Z(\tau^*) \mathbf{1}_{\{\tau^* \leq T\}} \middle| \mathcal{G}(t) \right] \\
 &+ \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T \mathbf{1}_{\{\tau^* > u\}} \exp \left(- \int_t^u r(s) ds \right) c(u) du \middle| \mathcal{G}(t) \right]. \quad (\text{II.1.2})
 \end{aligned}$$

The formula (II.1.2) is finally more useful for the modeling because the components are isolated.

Having stated the credit risk components and the pricing formula, it remains to specify the firm-value processes X and the default-free interest-rate process r to specialize the model. In the most commonly used models, Black-Scholes (GBM) dynamics is chosen for X and r is constant (see the Section II.1.4.1 for details of the Merton model). Clearly, one could instead of the 'limited' GBM use a more sophisticated model such as Heston's Stochastic Volatility (Heston, 1993), Variance Gamma (Madan et al., 1998), (Madan and Seneta, 1990), VGSA (Carr et al., 2003), CGMY (Carr et al., 2002), Bates (Bates, 1996)³³ or others, yet with their increasing complexity these are more difficult to calibrate. Because we can not map the asset process X to any real-world process and there is basically only one variable (firm's equity) that serves for the calibration, it would be rather a delicate task to apply one of the more advanced model in the modeling framework.³⁴ The choice of the dynamics of r is not so restrictive because to calibrate the dynamics of r to the market there is typically enough data. When choosing the model specification one should also take into account whether such a specification can yield analytical results. In the cases, when the model is complex, only a PDE (or PIDE) representation of the pricing equation exists but there are exceptions such as (Shimko et al., 1993) who found an analytical formula for Black-Scholes option price in which r is being driven by Vasicek's (Vasicek, 1977) interest-rate model. In the Section II.1.4.2 we present an analogy for the BS model with r subject to the arbitrage-free dynamics of (Hull and White, 1990).

³³see also the comparison of these models and their ability to fit the market quotes in (Kolman, 2014b) where the Fourier-Cosine expansion is used for calibration of the more complex models.

³⁴in the Section II.1.5 we show one possible but less traditional method of calibration that allows for fitting of more complex models.

To conclude this part, recall that although the very general valuation formula (II.1.2) typically simplifies when we adapt it to a particular model there may still appear significant complications resulting from the scarcity of the firm-value data and from other requirements on the model. Also, notice that if the default time τ^* is allowed to occur at any time (the so called *First passage time* models) until maturity T (that is $\mathcal{T} = [0, T]$) which is the real-world case, an additional dimension of complexity is present. Although this was addressed e.g. by (Black and Cox, 1976) or much later by (Leland, 1994) who both furthermore analyze the capital structure of the firm, the provided closed-form formulas are adapted to the Black-Scholes model with a constant risk-free rate as more general cases would yield non-analytical results. In turn, the most comprehensive specification of the pricing equation (II.1.2) is through PDE (or PIDE) which is the subject of the next section.

II.1.3 PDE framework for credit risk pricing

Valuation of (defaultable) claims may be conducted in many ways. If an analytical formula can be found it is arguably the most efficient method of valuation. However, in many cases one does not deal with problems that allow for analytical formulas. In these cases, we must proceed with some numerical method. To capture most of the phenomena in the market, it is appropriate to assume that the interest rates are stochastic and possibly correlated with other processes. Also, a company can generally default continuously in time, not only at some specific date which might be a very restrictive assumption. This possibility of default continuously in time would in the universe of the firm-value models require the existence of some default barrier. Many bonds, to the addition, involve embedded options. Combination of all these requirements demands a very robust and flexible numerical approach. One can automatically dismiss analytical solutions, tree models and transform methods. What remains is the Monte Carlo (MC) simulation and PDEs. Because MC methods are relatively slow, and do not yield unique results, these will also be dismissed from the selection. Eventually, we end up with PDEs which not only provide a very general environment, yet also provide easily reproducible results and also are faster than MC methods if the scheme is suitably optimized for the performance.

We assume that the assets X and the risk-free rate r are stochastic variables and follow Itô's dynamics. This clearly has an impact on the defaultable claim V which is subject to the default risk and also is sensitive to interest rates. To eventually find the main pricing PDE for V , the pricing problem V needs to be expressed as a suitably formed expectation which allows for the Feynman-Kac representation for V .

Our main objective is to reformulate (II.1.2) such that the 'suitable' conditional expectation is formed and Feynman-Kac can be applied. This is not a straightforward task because (II.1.2) involves various elements that are non-trivial. The approach we will be following is based on a *specialization* of the formula (II.1.2). We will also cover only the

pre-default³⁵ cases, that is $V(t, T)$ on the set $\{\tau^* > t\}$ only. Furthermore, we postulated earlier that the post-default recovery is *either* determined by δ , i.e. recovery settlement at maturity T regardless of the exact value of τ^* or by Z settled at $\tau^* \leq T$. We will label these two cases V_δ, V_Z , respectively.

Therefore, V_δ (V on the set $\{\tau^* > t\}$ with $Z \equiv 0$) equals

$$\begin{aligned} & V_\delta(t, T) \\ = & \mathbb{E}^\mathbb{Q} \left[\exp \left(- \int_t^{\tau^*} r(u) du \right) B(\tau^*, T) \delta \mathbf{1}_{\{\tau^* \leq T\}} + \exp \left(- \int_t^T r(u) du \right) g \mathbf{1}_{\{\tau^* > T\}} \middle| \mathcal{G}(t) \right] \\ + & \mathbb{E}^\mathbb{Q} \left[\int_t^T \mathbf{1}_{\{\tau^* > u\}} \exp \left(- \int_t^u r(s) ds \right) c(u) du \middle| \mathcal{G}(t) \right]. \end{aligned} \quad (\text{II.1.3})$$

This expression can be explained in terms of a simple logic. First assume away the 'coupon integral'. Then the expression captures two situations. (i) If the default has *not* occurred until the maturity T , the claimholder receives the T -terminal settlement g , whose present value is g times the discounting integral. (ii) When there *is* a default until T , then one receives at the default time τ^* discounted value of the recovery claim δ ³⁶ (which is to be received a.s. at T). The integral from t to τ^* then represents the present value of this 'rebate' received at τ^* as seen from the valuation time t . Comparing this with a stock option valuation, the payoff g is the payoff of the option if the stock does not hit the knock-out (default) barrier. If the stock hits the knock-out barrier at time τ^* , the option holder receives at τ^* a rebate R , in this defaultable claim case $R(\tau^*) = B(\tau^*, T)\delta$. To arrange the expression (II.1.3) for V_δ such that Feynman-Kac can be invoked (cf. Theorem I.1.5), we rewrite it as

$$V_\delta(t, T) = \mathbb{E}^\mathbb{Q} \left[\exp \left(- \int_t^T r(u) du \right) g + \int_t^T \exp \left(- \int_0^u r(s) ds \right) \tilde{c}(u) du \middle| \mathcal{G}(t) \right], \quad (\text{II.1.4})$$

where the valuation problem can be subject to barriers and $\tilde{c}(u) = \mathbf{1}_{\{\tau^* > u\}} c(u)$. If the default barrier h (if defined) is hit at time τ^* (i.e. at default), then the claimholder receives a rebate $R(\tau^*)$ equal to

$$R(\tau^*) = B(\tau^*, T)\delta.$$

Notice that we derived (II.1.3) (and hence also (II.1.4)) from (II.1.2) (besides other things)

³⁵pre-default in the sense of alive, not necessarily very close to default.

³⁶although one of the model's assumptions was that after default the bondholder receives δ at T , here we postulate that he receives $B(\tau^*, T)\delta$ at τ^* . This is not a breach of the assumption as in the risk-neutral world one is indifferent whether to receive δ at T or $B(\tau^*, T)\delta$ at τ^* since δ is a *sure* payoff at T , conditionally on $\tau^* \leq T$.

using the fact that

$$\begin{aligned}
& \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T r(u) du \right) \delta \mathbf{1}_{\{\tau^* \leq T\}} \middle| \mathcal{G}(t) \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \left\{ \int_t^{\tau^*} r(u) du + \int_{\tau^*}^T r(u) du \right\} \right) \delta \mathbf{1}_{\{\tau^* \leq T\}} \middle| \mathcal{G}(t) \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{\tau^*} r(u) du \right) \exp \left(- \int_{\tau^*}^T r(u) du \right) \delta \mathbf{1}_{\{\tau^* \leq T\}} \middle| \mathcal{G}(t) \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{\tau^*} r(u) du \right) \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_{\tau^*}^T r(u) du \right) \middle| \mathcal{G}(\tau^*) \right] \delta \mathbf{1}_{\{\tau^* \leq T\}} \middle| \mathcal{G}(t) \right] \\
&= \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{\tau^*} r(u) du \right) B(\tau^*, T) \delta \mathbf{1}_{\{\tau^* \leq T\}} \middle| \mathcal{G}(t) \right].
\end{aligned}$$

In the sequence of expressions above we firstly decomposed the discounting integral into two pre- and post- default periods (t, τ^*) , $[\tau^*, T]$, respectively. Then, we used the property of the iterated expectations³⁷ and expressed the post-default discount factor as a risk-free bond value $B(\tau^*, T)$.

Let us now focus on the V_Z case, with the recovery $Z(\tau^*)$ paid at τ^* . This means we again use the expression (II.1.2) for V and specialize it to only pre-default value (enforcing the pre-default set $\{\tau^* > t\}$) and without the recovery payment δ paid at maturity T , yet including Z . This gives

$$\begin{aligned}
& V_Z(t, T) \\
&= \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{\tau^*} r(u) du \right) Z(\tau^*) \mathbf{1}_{\{\tau^* \leq T\}} + \exp \left(- \int_t^T r(u) du \right) g \mathbf{1}_{\{\tau^* > T\}} \middle| \mathcal{G}(t) \right] \\
&+ \mathbb{E}^{\mathbb{Q}} \left[\int_t^T \mathbf{1}_{\{\tau^* > u\}} \exp \left(- \int_t^u r(s) ds \right) c(u) du \middle| \mathcal{G}(t) \right]. \tag{II.1.5}
\end{aligned}$$

In spite of the different assumptions about the recovery policy, this is surprisingly the very same expression as for the V_δ case (cf. (II.1.3)), with the only exception that instead of $B(\tau^*, T)\delta$ here we have $Z(\tau^*)$. We conclude that the rebate function R for V_Z equals

$$R(\tau^*) = Z(\tau^*).$$

Therefore, for V_Z we necessarily have the same PDE as for V_δ with the only difference in the specification of the rebate function R . The identity of the PDEs for V_δ and V_Z is a particularly beneficial observation since both cases V_δ and V_Z can be handled within the *same PDE framework* by only changing the choice of R .

³⁷sometimes also paraphrased as 'the coarser algebra wins'.

Because both claims V_δ, V_Z (equations (II.1.4) and (II.1.5), respectively) differ only in the rebate function R , we can finally form a universal 'suitable' expectation for both these claims (denoted by V) and we have

$$V(t, T) = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T r(u) du \right) g + \int_t^T \exp \left(- \int_t^u r(s) ds \right) \tilde{c}(u) du \middle| \mathcal{G}(t) \right], \quad (\text{II.1.6})$$

where $\tilde{c}(u) = c(u)\mathbf{1}_{\{\tau^* > u\}}$ and subject to the *barrier rebate* function

$$R(\tau^*) = \begin{cases} B(\tau^*, T)\delta & \text{for } \delta\text{-recovery} \\ Z(\tau^*) & \text{for } Z\text{-recovery} \end{cases}, \quad (\text{II.1.7})$$

given τ^* is subject to some barrier h .

Since the claim $V(t, T)$ is actually $V(t, T, r, X)$, where r, X are Itô's processes, and given the particular form of the expression (II.1.6) for V , it is clear from the Theorem I.1.5 that Feynman-Kac applied and so there exists a particular pricing PDE for V . We postulate the following key result.

Theorem II.1.1 (Fundamental pricing PDE (firm-value models))

Let us assume that $x_1 = r$ and $x_2 = X$ (compactly $x = (x_1, x_2)$) follow

$$dx_i(t) = \mu_i(t, x)dt + \sigma_i(t, x)dW_i^{\mathbb{Q}}(t), i = 1, 2,$$

given $dW_1^{\mathbb{Q}}(t)dW_2^{\mathbb{Q}}(t) = \rho dt$. Then the pre-default value of the claim V of the kind (II.1.6) solves

$$\frac{\partial V}{\partial t} + \sum_{i=1}^2 \mu_i(t, x) \frac{\partial V}{\partial x_i} + \frac{1}{2} \sum_{i=1}^2 \sigma_i^2(t, x) \frac{\partial^2 V}{\partial x_i^2} + \rho \sigma_1(t, x) \sigma_2(t, x) \frac{\partial^2 V}{\partial x_1 \partial x_2} + \tilde{c}(t, x) - x_1 V = 0, \quad (\text{II.1.8})$$

with the terminal condition $V(T, x) = g(x)$. The value of the claim V is further subject to a barrier h where the recovery rebate $R(\tau^*)$ is defined in (II.1.7).

Remarks. (i) Although it is assumed that there are two stochastic factors $x_1 = r, x_2 = X$ this can be any number of factors yet the PDE would be much more complex and difficult to maintain in a numerical scheme. (i) In the firm-value modeling environment, the default barrier h is typically present. This would have to be reflected by setting appropriate boundary conditions which commonly be of the Dirichlet type.

The presented PDE is very general and together with the specification of the boundary conditions allows for variety of settings, thus it is suitable for a wide class of defaultable claims valuation problems where x_1, x_2 follow Itô's processes. Adding *pure jumps*³⁸ would imply to use PIDE³⁹ specification instead. Also, not necessarily we need to stick with interest-rate models that allow for analytical solutions for bond prices but less traditional models can be used. In fact when both x_1, x_2 are stochastic processes, things

³⁸Wiener process is regarded as a jump process in spite of its perfect continuity. It is however *not a pure jump process*.

³⁹partial integro-differential equation.

get much more complicated because there is a two-dimensional PDE in place. Such a two-dimensional PDE can efficiently be solved using e.g. the ADI⁴⁰ (see e.g. (Craig and Sneyd, 1988) or (Hirsa, 2016, Section 4.7)). One technical complication for ADI, however, arises when $\rho \neq 0$. In such a case, the PDE can not be simply evaluated in the ADI scheme because in the ADI-based methods, derivatives are taken into x_1 or x_2 direction in a single step but not into both directions at once, yet when $\rho \neq 0$ this is required.⁴¹ To overcome these complications, we suggest to conduct a PDE *orthogonalization* which transforms a PDE involving mixed derivatives such as (II.1.8) into a PDE with no mixed derivatives. For details see (Andersen and Piterbarg, 2010, Section 2.11.1). We also investigated the ADI-based methods in (Kolman, 2015b).

We also note that the PDE (II.1.8) can also support pricing problems with embedded options (or convertible bonds). These can, for example, be either linked to, say, upper barrier,⁴² or more commonly there could be some pre-specified set of times $\tilde{\mathcal{T}}$ at which we compare the value V on the grid at every $\tilde{\tau} \in \tilde{\mathcal{T}}$ with the strike price of the embedded option. This would then correspond to a *Bermudan* claim valuation and in the limit, to the American claim valuation although this is unusual in the fixed-income market. Several applications of this PDE can be found in (Kolman, 2016) where it is also discussed how the PDE differs from the 'equivalent' PDE in the reduced-form framework.

Also, as stated above, the promised dividends are represented by a coupon process c which is by assumption a *continuously* paid coupon rate. This is hardly compatible with the market practice, since the coupons are typically paid quarterly or (semi)annually. This difference in the cash-flow profile could potentially impair the pricing⁴³ as the continuously paid coupons represent only a *proxy* for the real-world market. This approximation is, however, possible to be overcome. Since a coupon payment can be seen as a zero-bond with zero recovery a coupon-bearing bond can under some circumstances be replicated by a zero-coupon bond with some recovery rule plus a portfolio of these zero-recovery zero-bonds whose face value is scaled by the coupon payment. One would therefore be challenged to solve multiple PDEs, one for the 'main' zero-bond plus one PDE per a coupon payment.⁴⁴

⁴⁰Alternating-direction implicit.

⁴¹ $\rho \neq 0$ ensures the existence of a mixed derivative given the diffusion components are non-zero.

⁴²for example the firm-asset value x_2 might reach some threshold where it could be beneficial withdraw (call) the bond and issue a new bond with a lower coupon. Clearly also the same might hold for some low level of the risk-free rate $r = x_1$.

⁴³the inaccuracy is a two-fold problem. (i) the present value of a cash-flow paid continuously over time differs from the same cash-flow (accumulated and) paid at one time-point, and more importantly (ii) the incidence of default is greater in the case of discretely paid coupons. If the coupons are paid continuously, the accrued interest is immediately paid out, whence in the discrete case, the accrued interest accumulates for a certain period. If the default occurs exactly before the coupon payment the coupon losses in the continuous case are negligible relative to the discrete case where the claimholder loses almost the whole coupon payment.

⁴⁴such a valuation approach is generally *not fully* compatible with all the proposed methodologies and rather serves for plain pricing problems. For example, one can not use this approach for pricing bonds with embedded options because the coupon payments are now separated from the valuation of the face. However, for the decision whether to exercise or not, one needs to take the total bond value at the call/put date $\tilde{\tau}$ into consideration, yet the coupons are now distributed to multiple PDEs.

Finally, let us end this section with two useful special cases which are widely known. If we in the PDE (II.1.8) eliminate the stochasticity of the risk-free rate by setting $\mu_1 \equiv 0, \sigma_1 \equiv 0, \tilde{c} \equiv 0$, relabel x_1 to r and x_2 to S , define $\mu_2(t, x_1, x_2) = rS(t), \sigma_2(t, x_2) = \sigma S(t)$ and assume $V(T) = V(T, S(T))$, then the PDE for V satisfies

$$\frac{\partial V}{\partial t} + rS \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} - rV = 0,$$

which is actually the Black-Scholes PDE.

Analogously, eliminating the effect of the firm-value variable x_2 by setting $\mu_2(t, x_2) \equiv 0, \sigma_2(t, x_2) \equiv 0$, relabelling x_1 to r , simplifying $\mu_1(t, x_1, x_2) = \mu(t, r)$, $\sigma_1(t, x_1, x_2) = \sigma(t, r)$ and setting $\tilde{c} \equiv 0$, gives for some function $V(t, T, r)$ (with $V(T, T, r) = 1$) the PDE

$$\frac{\partial V}{\partial t} + \mu(t, r) \frac{\partial V}{\partial r} + \frac{1}{2} \sigma^2(t, r) \frac{\partial^2 V}{\partial r^2} - rV = 0.$$

This is the risk-free zero bond $V(t, T) \equiv B(t, T)$ pricing PDE (see (Björk, 2009, Proposition 23.2) where this equation is known as the 'Term structure equation'). These two marginal cases neatly illustrate the general connection of the PDE (II.1.8) with both the capital market and the fixed-income market.

II.1.4 Models with closed-forms

In the previous sections we have thoroughly discussed a *general* firm-value framework. Although this yielded powerful, completely general pricing formula for defaultable claims, in practice we usually work with some well-established closed-form models. We thus find it useful to discuss them as these will be later applied in an analysis (Section II.1.5).

II.1.4.1 Merton's model

Merton's model (Merton, 1974) is the pioneer among the large class of structural models. It is based on Black-Scholes model for stock options, with all its assumptions and imperfections. The main benefits of the Merton model are the straightforward calibration and the analytical results. Probably the biggest drawback of Merton's model is the very limiting assumption about the default time τ^* which is only allowed to occur at maturity T of a hypothetical zero-bond issued by the company. The Merton's model is fundamentally dependent on the following two assumptions:

Assumption II.1.1 (Merton's model)

- *the company is funded only by the equity E and a zero-bond with face value K maturing at T ,*
- *default can occur only at maturity T of the zero-bond.*

We are now in the position to find an option pricing analogy in Merton's approach. The option analogy stems from the possible T -payoffs for (i) stockholders and (ii) bondholders.

Let us analyze the case (i) – the stockholders' payoff:

- if there are sufficient assets X by time T to pay off the debt liability K , then the stockholders receive the excessive value of the assets over the debt, i.e. $E(T) = X(T) - K$,
- if there are by time T insufficient assets X to pay the debt K , then the stockholders receive 0,
- collecting both scenarios into one expression gives the stockholders' payoff $E(T) = \max[X(T) - K, 0]$.

Proceeding similarly for (ii) – the bondholders' payoff:

- if there are sufficient assets X by time T to pay off the debt liability K , then bondholders receive the face value K *in full*,
- if there are by time T insufficient assets X to pay the debt K , then the bondholders receive only the smaller of the two quantities, thus X ,
- collecting both scenarios into one expression gives the bondholders' payoff $V(T) = \min[X(T), K] = K - \max[K - X(T), 0]$.⁴⁵

After summarizing, we have:

- the value of the *equity* E is equal to the price of a *long call* on the company *assets* X with a strike K equal to the face value of the debt due at T ,
- the payoff $V(T)$ to the bondholders is equal to a *long* position on the *debt* of face K and a *short* position in a *put* option with strike K on company's *assets* X . Thus, the present value of V is equal to the present value of an expected value of such a portfolio.

Let us denote $C(t, T)$, $P(t, T)$ the value of a call and a put option, respectively, on the assets X of a given firm. Then, at time t , the present value of the expected payoffs for the stockholders and bondholders, respectively, equals

$$E(t) = C(t, T) \quad (\text{II.1.9})$$

$$V(t, T) = Ke^{-r(T-t)} - P(t, T). \quad (\text{II.1.10})$$

Finally, because the assets $X(t)$ are represented by the present value of the equity $E(t)$ and the present value of the debt $V(t, T)$, the following identity must hold for the total assets

$$X(t) = E(t) + V(t, T). \quad (\text{II.1.11})$$

⁴⁵this is achieved by the usual min-max transformation:

$$\begin{aligned} \max[x, y] &= x\mathbf{1}_{\{x \geq y\}} + y\mathbf{1}_{\{x < y\}} \\ \min[x, y] &= x\mathbf{1}_{\{x < y\}} + y\mathbf{1}_{\{x \geq y\}} = x(1 - \mathbf{1}_{\{x \geq y\}}) + y(1 - \mathbf{1}_{\{x < y\}}) \\ &= x + y - x\mathbf{1}_{\{x \geq y\}} - y\mathbf{1}_{\{x < y\}} = x + y - \max[x, y] \\ &= x + y - (\max[0, y - x] + x) \\ &= y - \max[0, y - x]. \end{aligned}$$

Using the theory developed in the Section II.1.2 it is easy to find a particular pricing formula for the defaultable claim $V(t, T)$ on the set $\{\tau^* > t\}$. The result is stated in the following theorem.

Theorem II.1.2 (Pre–default value of debt in the Merton model)

In view of the generalized risk-neutral pricing formula defaultable claims (II.1.2), the value of company’s T –debt of a face value K prior to default is equal to

$$V(t, T) = B(t, T)\mathbb{E}^{\mathbb{Q}} [X(T)\mathbf{1}_{\{\tau^*=T\}} | \mathcal{G}(t)] + KB(t, T)\mathbb{E}^{\mathbb{Q}} [\mathbf{1}_{\{\tau^*>T\}} | \mathcal{G}(t)],$$

where $\tau^* = T\mathbf{1}_{\{X(T)<K\}} + \infty\mathbf{1}_{\{X(T)\geq K\}}$.⁴⁶

Proof. In (II.1.2) first fix the pre-default set $\{\tau^* > t\}$, and replace all discounting integrals with $B(t, T)$ due to a constant interest rate r . Since the debt of a face value K represents a zero-coupon bond then the coupon rate is trivially $c = 0$. Because further the default can *not* occur at a random time prior the maturity T , we can set the ‘recovery claim’ Z to 0. We are left with only two terms. The first one assumes a (possibly random) recovery payoff δ at T the when default occurs prior or at T and the other term ensures payment of K if there is no default until T (including T itself). It thus only remains to further narrow the default time condition and the recovery payment δ . Thus, we specify the default time τ^* as $\tau^* = T\mathbf{1}_{\{X(T)<K\}} + \infty\mathbf{1}_{\{X(T)\geq K\}}$ in order to ensure that default can occur only at the maturity T . Finally, since the default bond pays $X(T)$ upon default τ^* and the default is only allowed to occur at T , then $\delta = X(T)$. \square

Also, the defaultable claim V in terms of the proposed PDE framework (see the Section II.1.3) has the following simple specification.

Theorem II.1.3 (PDE for the debt value in the Merton model)

The debt $V = V(t, T)$ of a face value K and maturing at T , prior to default is a solution to the PDE

$$\begin{aligned} \frac{\partial V}{\partial t} + \left(r - q - \frac{1}{2}\sigma_x^2 \right) \frac{\partial V}{\partial x} + \frac{1}{2}\sigma_x^2 \frac{\partial^2 V}{\partial x^2} - rV &= 0 \\ V(T, T, x) &= \min[e^x, K], \end{aligned}$$

where $x = \ln X$ is the log-asset value process, σ_x is the asset volatility and q is the asset payoff rate.

Proof. The result comes directly from the Theorem II.1.1 just by specifying the T –payoff $V(T, T) = \min[X(T), K]$ and setting the risk-free rate to a constant r . \square

A more traditional approach to analysis of the Merton model is, however, typically based on the nice analytical properties of the terms involved in (II.1.9) and (II.1.10) in conjunction with (II.1.11) since these can easily be solved in the Black-Scholes framework. We omit the derivation and present directly the well-known formulas for the call and the

⁴⁶in this particular case we assume $\infty \times 0 = \infty$.

put option which are in this case

$$\begin{aligned}
C(t, T) &= X(t)e^{-q(T-t)}N(d_1) - Ke^{-r(T-t)}N(d_2) & (\text{II.1.12}) \\
P(t, T) &= Ke^{-r(T-t)}N(-d_2) - X(t)e^{-q(T-t)}N(-d_1) \\
d_1 &= \frac{\ln\left(\frac{X(t)}{K}\right) + \left(r - q + \frac{1}{2}\sigma_x^2\right)(T-t)}{\sigma_x\sqrt{T-t}} \\
d_2 &= d_1 - \sigma_x\sqrt{T-t}.
\end{aligned}$$

Analyzing the yield spread $s(t, T) = y(t, T) - r(t, T) = y - r$, we shall firstly note that

$$\begin{aligned}
y &= \frac{\ln K - \ln V(t, T)}{T-t} = \frac{\ln K - \ln(Ke^{-r(T-t)} - P(t, T))}{T-t} \\
&= \frac{\ln K - \ln(Ke^{-r(T-t)}(1 - N(-d_2)) + X(t)e^{-q(T-t)}N(-d_1))}{T-t},
\end{aligned}$$

and so

$$s(t, T) = \frac{\ln K - \ln(Ke^{-r(T-t)}(1 - N(-d_2)) + X(t)e^{-q(T-t)}N(-d_1))}{T-t} - r,$$

is the closed-form model credit spread formula.

What has not been discussed so far is the **default threshold** K . This quantity (together with the effective maturity T) is a very delicate, partially subjective input to the model. In practice, companies are hardly ever financed by a single zero-bond but have a rather more diverse debt structure which is furthermore due in various time horizons not at a single date T . Here, to set K correctly one is thus rather left to using own experience (observations) of the default cases. For example, the business-wise successful product KMV (Crosbie and Bohn, 2003) (or alternatively (Crouhy, 2014)) which is based on Merton's model, revealed that companies usually default when the assets X fall somewhere between the short-term liabilities plus 1/2 of the long-term liabilities. The time T is generally approximated by weighted maturities (or duration) of the liabilities.

Next, we shall discuss the **calibration**. In Merton model, calibration means finding the pair of inputs $(X(t), \sigma_x)$ as these are not observable on the market. In fact, there are two common methods of finding $(X(t), \sigma_x)$ (see (Cossin, 2001, Section 3.6) or (Loffler, 2007, Section 2)). The first approach is rather 'statistical' as it requires creating a time series of assets (PV of tradeable debt + PV of non-tradeable debt + equity), where the PV of the non-tradeable debt is calculated using the same book-to-market ratio as the tradeable debt. Then, the assets X and their volatility σ_x can directly be inferred from the established time series. The second and maybe more typical approach is through the connection of E and the hypothetical call on assets C by (II.1.9). However, two unknowns require two equations. The second equation is obtained as follows. Since a call option is a function of the Itô's process X , the differential dC can be obtained by the Itô's lemma (see Corollary I.1.1). Because furthermore also the equity E follows Itô's process (GBM) and the value of E is equal to $C(\cdot, T)$ their differentials must coincide. Comparison of the

diffusion terms then gives the second expression connecting the asset volatility σ_x and the equity volatility σ_E .

To see this, Itô's lemma for $dC(t, T) = df(t, X(t))$ gives

$$\begin{aligned} dC(t, T) &= \frac{\partial C}{\partial t} dt + \frac{\partial C}{\partial X} dX(t) + \frac{1}{2} \frac{\partial^2 C}{\partial X^2} (dX(t))^2 \\ &= (\dots) dt + \frac{\partial C}{\partial X} X(t) \sigma_x dW_x^{\mathbb{Q}}. \end{aligned}$$

Because the equity E is assumed to follow a GBM, we in turn have (under \mathbb{Q})

$$dE(t) = (\dots) dt + \sigma_E E(t) dW_x^{\mathbb{Q}}.$$

Comparison of the diffusion terms in $dC(t, T)$ and $dE(t)$ gives

$$\sigma_x = \sigma_E \frac{E(t)}{X(t)} \frac{1}{N(d_1)}. \quad (\text{II.1.13})$$

As discussed above, the second equation is then (II.1.9), that is

$$E(t) = C(t, T), \quad (\text{II.1.14})$$

with $C(t, T)$ defined in (II.1.12). The system (II.1.13) and (II.1.14) then needs to be solved simultaneously for $(X(t), \sigma_x)$.

We, however, do not recommend this approach to calibration for the *pricing* purposes. A phenomenon of producing lower than observed credit spreads, more colloquially known as the 'credit spread puzzle', discussed in the recent papers (Feldhütter and Schaefer, 2015), (Huang and Huang, 2012) or previously in (Chen et al., 2009), usually appears in the Merton model. The basic essence of this market inconsistency stems from the equation (II.1.13). That equation implies that σ_x is typically strictly lower than σ_E which perfectly makes sense from the balance sheet point of view⁴⁷ but may not be appropriate for the credit spread analysis. In the Figure 1 we show a comparison of implied spreads that we obtained for three levels of leverage and various equity volatilities σ_E by taking these values (leverage, σ_E) as inputs, running the 'classical' calibration procedure above to estimate $(X(t), \sigma_x)$ and computed the 5Y spreads from these quantities. What is surprising is that given some value of σ_E , the highly leveraged company exhibits *lower* spread than the mid-leveraged company. The second clear message from this figure is that if the equity volatility σ_E is lower than 30% in this particular setup, there is almost no credit spread regardless of the leverage. This is against the empirical observations on the market.

As the standard calibration exhibit deficiencies, in the Section II.1.5 dedicated to the empirical analysis, we compare the credit spreads generated by this standard calibration method with spreads obtained through a more elaborate method based on option quotes.

II.1.4.2 Merton's model with stochastic r (Hull-White)

The minor limiting issue in the Merton model is its inability to capture stochasticity of the interest rates. Also, in the real world, there is some risk-free term structure observed

⁴⁷since the assets X are represented by the debt K with no volatility and the equity E , then X shall obviously have smaller volatility than the equity E .

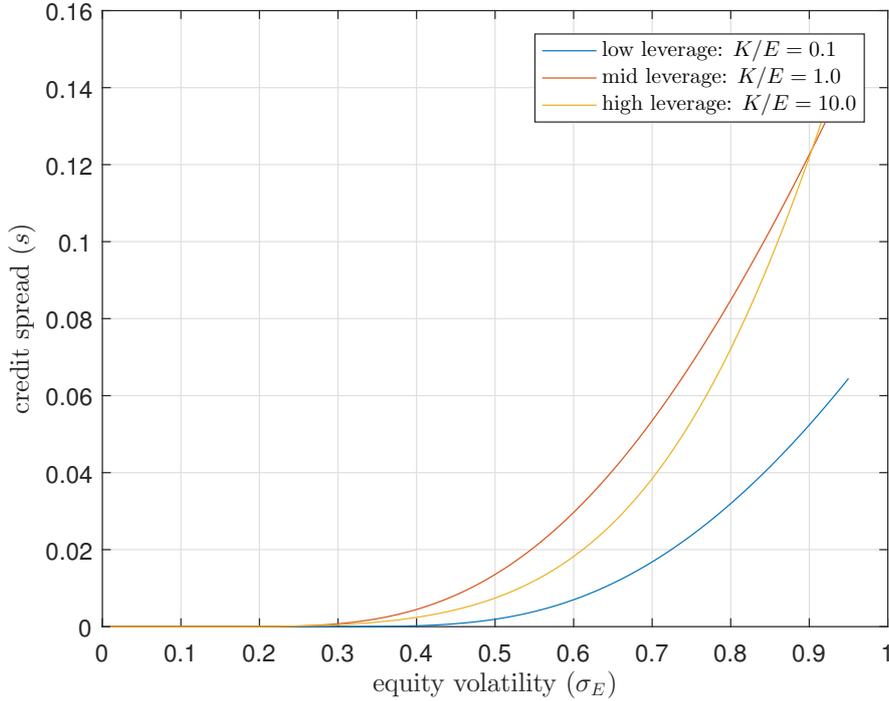


Figure 1: Strange behaviour of the credit spread in the Merton model when calibrated using the standard method for three levels of the financial leverage.

on the market while the Merton model assumes $r(t) = r$, that is the yield curve is flat. It would be arguably more appropriate to use some sort of stochastic model for the risk-free rate and combine it with the Merton framework. (Shimko et al., 1993) has shown this is possible when Merton model is combined with the Vasicek interest-rate model. We extend this to the Hull-White⁴⁸ (Hull and White, 1990) model which is perhaps more suitable model for the pricing than Vasicek because Hull-White is arbitrage-free model while Vasicek is an 'equilibrium' model.

The idea is to specify the \mathbb{Q} dynamics of r and X as

$$\begin{aligned} dr(t) &= \kappa(\theta(t) - r(t))dt + \sigma_r dW_1^{\mathbb{Q}}(t) \\ dX(t) &= (r(t) - q)X(t)dt + \sigma_x X(t)(\rho dW_1^{\mathbb{Q}}(t) + \sqrt{1 - \rho^2} dW_2^{\mathbb{Q}}(t)), \end{aligned}$$

where the mean-reversion level θ is chosen as

$$\theta(t) = \frac{1}{\kappa} \frac{\partial f(0, t)}{\partial t} + f(0, t) + \sigma_r^2 \frac{1 - e^{-2\kappa t}}{2\kappa^2},$$

to ensure a perfect fit to the observed yield curve.

⁴⁸the result is actually valid for any HJM-based model. Here, we stick with Hull-White.

Based on this specification the solution for the call and put option, respectively, is

$$\begin{aligned}
C(t, T) &= X(t)e^{-q(T-t)}N(d_1) - KB(t, T)N(d_2) \\
P(t, T) &= KB(t, T)N(-d_2) - X(t)e^{-q(T-t)}N(-d_1) \\
d_1 &= \frac{\ln\left(\frac{X(t)}{KB(t, T)}\right) + \frac{1}{2}v_{\text{HW}}(t, T) - q(T-t)}{\sqrt{v_{\text{HW}}(t, T)}} \\
d_2 &= d_1 - \sqrt{v_{\text{HW}}(t, T)} \\
v_{\text{HW}}(t, T) &= \frac{\sigma_r^2(2\kappa(T-t) + 4e^{-\kappa(T-t)} - e^{-2\kappa(T-t)} - 3)}{2\kappa^3} \\
&+ \frac{\sigma_x(\kappa^2\sigma_x(T-t) + 2\rho\sigma_r(\kappa(T-t) + e^{-\kappa(T-t)} - 1))}{\kappa^2}. \quad (\text{II.1.15})
\end{aligned}$$

Proof. See the Appendix E. □

The analysis within this stochastic model is essentially similar to the analysis in the classical Merton model, yet now we have the extra element - the stochastic r whose dynamics needs to be appropriately calibrated to fit the model to the market. We now discuss this calibration issue in a greater detail.

Calibration of the stochastic model. The calibration essentially consists of three steps. Firstly, calibrate the Hull-White model to the prices of interest-rate derivatives (mostly caps/floors or swaptions⁴⁹), see for example (Málek, 2005), (Brigo and Mercurio, 2006), or (Andersen and Piterbarg, 2010). Then, estimate the correlation ρ either by computing the correlation of returns of the time series r and E (the proxy method) or more appropriately apply the custom-developed PDE calibration method outlined in the Appendix D by jointly considering the differentials dE , dr and fixing all parameters except ρ . Because the same Wiener process drives E and X and since furthermore the correlation ρ is measure-invariant (see (Shreve, 2004, p. 258)), then ρ estimated this way is the best possible point estimate of the correlation of the Wiener processes in r and X . Having all the parameters of the r -process and ρ , it remains to estimate $(X(0), \sigma_x)$ the same way as in the standard Merton model with the terms d_1 , $C(t, T)$ stated in this section which are customized for the stochastic rate r .

II.1.4.3 Zhou's jump diffusion model

The perfect predictability of the default time τ^* in Merton's model might be seen as a drawback because default of a company might occur also randomly. The Merton model, in fact, implicitly assumes that the markets are perfectly efficient because the default time τ^* does not surprise the investors. This is, however, not true in practice, as companies often default 'unexpectedly' as there is some informational asymmetry present in the markets. For an exhaustive and very informative explanation of the information flow in credit models, see (Jarrow and Protter, 2004), who link the (in)accessibility of the default time τ^* to the empirical examples. In turn, to relax the assumption of the efficient markets, the perfect accessibility of the default time τ^* needs to be weakened. A natural model-wise

⁴⁹see the Appendix A.3–A.4 for a closer discussion.

way to do this is to introduce jumps into the asset process X and this exactly has been addressed by (Zhou, 1997). Zhou assumes that the assets X under \mathbb{Q} follow the JDE

$$dX(t) = (r - q - \tilde{\lambda}\tilde{\beta})X(t)dt + \sigma_x X(t)dW^{\mathbb{Q}}(t) + X(t)(J(t) - 1)d\tilde{N}(t),$$

where $\ln(J(t)) \sim N(\mu_J, \sigma_J^2)$ is an independent jump size and \tilde{N} is a Poisson process under \mathbb{Q} . With a slight abuse of notation, denoting $J_i = \left\{ J(t) : \tilde{N}(t) = i \mid \tilde{N}(t-) = i - 1 \right\}$ (that is the size of i th jump) it can be verified that the solution of such JDE is

$$X(t) = X(0) \exp \left(\left(r - q - \frac{1}{2}\sigma_x^2 - \tilde{\lambda}\tilde{\beta} \right) t + \sigma_x W^{\mathbb{Q}}(t) \right) \prod_{i=1}^{\tilde{N}(t)} J_i.$$

Because the pure jump component $X(t)(J(t) - 1)d\tilde{N}(t)$ is in general not a (\mathbb{Q} -) martingale the compensator term $X(t)\tilde{\lambda}\tilde{\beta}dt$ is present to clear away any systematic jump movement and ensure the local rate of return under \mathbb{Q} is $(r - q)$. Here, $\tilde{\beta}$ compensates the jump *size*, which is⁵⁰ $(J(t) - 1)$. Hence, we have

$$\tilde{\beta} = \mathbb{E}^{\mathbb{Q}} [J(t) - 1] = \int_{-\infty}^{\infty} e^{\mu_J + \sigma_J z} f_N(z) dz - 1 = \exp \left(\mu_J + \frac{1}{2}\sigma_J^2 \right) - 1.$$

The default time specification in Zhou's model⁵¹ remains the same as in the case of Merton, that is

$$\tau^* = T \mathbf{1}_{\{X(T) \leq K\}} + \infty \mathbf{1}_{\{X(T) > K\}},$$

and similar is also the option pricing analogy for the credit risk. That is the bondholders are long in the debt K and short in a put P on the company's assets X . The equity holders receive a hypothetical payoff equal to a long position in a call C with a strike K on the assets X . It is therefore very useful to have closed-form formulas for the put and call options. The solutions for call and put on the company's assets X read

$$\begin{aligned} C(t, T) &= X(t) e^{-q(T-t)} \sum_{i=0}^{\infty} e^{-\tilde{\lambda}(T-t)(\tilde{\beta}+1) + i\mu_J + \frac{1}{2}i\sigma_J^2} \frac{(\tilde{\lambda}(T-t))^i}{i!} N(d_1^{(i)}) \\ &\quad - K e^{-r(T-t)} \sum_{i=0}^{\infty} e^{-\tilde{\lambda}(T-t)} \frac{(\tilde{\lambda}(T-t))^i}{i!} N(d_2^{(i)}) \\ P(t, T) &= K e^{-r(T-t)} \sum_{i=0}^{\infty} e^{-\tilde{\lambda}(T-t)} \frac{(\tilde{\lambda}(T-t))^i}{i!} N(-d_2^{(i)}) \\ &\quad - X(t) e^{-q(T-t)} \sum_{i=0}^{\infty} e^{-\tilde{\lambda}(T-t)(\tilde{\beta}+1) + i\mu_J + \frac{1}{2}i\sigma_J^2} \frac{(\tilde{\lambda}(T-t))^i}{i!} N(-d_1^{(i)}), \end{aligned}$$

⁵⁰by this specification of the jump component if there is a jump at time t , then $X(t) = X(t-)J(t)$, that is the after-jump value of the assets $X(t)$ is *proportional* to $J(t)$.

⁵¹in fact the specification of the default time as $\tau^* = T \mathbf{1}_{\{X(T) \leq K\}} + \infty \mathbf{1}_{\{X(T) > K\}}$ is just a particular version of Zhou's model, who in general assumes default being triggered when a particular default barrier is hit.

where

$$\begin{aligned}
d_1^{(i)} &= \frac{\ln\left(\frac{X(t)}{K}\right) + \mu_i(T-t) + \sigma_i^2(T-t)}{\sigma_i(T-t)} \\
d_2^{(i)} &= d_1^{(i)} - \sigma_i(T-t) \\
\mu_i(\tau) &= \left(r - q - \frac{1}{2}\sigma_x^2 - \tilde{\lambda}\tilde{\beta}\right)\tau + i\mu_J \\
\sigma_i^2(\tau) &= \sigma_x^2\tau + i\sigma_J^2.
\end{aligned}$$

Although possible, we refrain from deriving a closed-form formula for the credit spread $s(t, T)$ because it is unnecessarily complicated and it can be computed just simply by using the general value-of-the-debt equation

$$V(t, T) = Ke^{-r(T-t)} - P(t, T) = Ke^{-(r+s(t, T))(T-t)},$$

and computing the spread $s(t, T)$ thus gives

$$s(t, T) = \frac{-\ln\left(e^{-r(T-t)} - \frac{P(t, T)}{K}\right)}{T-t} - r.$$

In spite of the richer specification of Zhou over the Merton's model, the practical complication with Zhou is how to calibrate it as it can not be done the same way as the Merton model due to the extra jump term. This, however, is not a limitation for the analysis in the Section II.1.5 as there we work with a special, quite general calibration method which is easily applicable to the Zhou model.

II.1.4.4 The Hestonian approach

Last but not least, we include the Heston model (Heston, 1993) specification into the list although it was originally not developed for the credit modeling. This is because in what follows we will use a general calibration method that is not limiting in the model choice and given an option pricing model, we can always recover a credit spread. As we will later calibrate the models to options that are, in general, priced via implied volatility with some volatility skew/smile, we find it natural cover the Heston model too as it was primarily designed to capture the volatility skew/smile on the market. This section thus concentrates on the model itself and its technical pitfalls, rather than on some reasoning for credit modeling.

According to Heston, the underlying (which we regard as the assets X) is subject to the following \mathbb{Q} -dynamics:

$$\begin{aligned}
dX(t) &= (r - q)X(t)dt + \sqrt{v(t)}X(t)dW_x^{\mathbb{Q}}(t) \\
dv(t) &= \kappa(\theta - v(t))dt + \sigma dW_v^{\mathbb{Q}}(t). \\
dW_x^{\mathbb{Q}}(t)dW_v^{\mathbb{Q}}(t) &= \rho dt.
\end{aligned}$$

Therefore the assets X follow a GBM with stochastic volatility $\sqrt{v(t)}$ and the Wiener processes of the assets X and the variance v are instantaneously correlated. Our objective,

however, is not to discuss the Heston model as such but rather to provide details about the implementation which is suggested for the credit risk modeling to minimize the numerical errors. An exhaustive discussion of the Heston model is provided by (Rouah, 2013).

In fact, the Heston model can be implemented in many ways. The most natural implementation is provided in the original article (Heston, 1993). This, however, is not regarded as being the best possible implementation due to some empirical integration issues, although it is mathematically correct. The main difficulty with Heston is that the PDFs of the random variable $X(T)$ and even of $x(T) = \ln X(T)$ do not exist and we are only left with semi-analytical solutions based on the characteristic functions (CFs) $\varphi_j, j = 1, 2$ of $x(T)$ which have analytical forms. There are two known versions of the CFs which are both valid yet they imply some differences when one is about to integrate them. We will stick with the 'alternative' version of the CF, strongly suggested by (Albrecher et al., 2007) for a better behaviour during a numerical integration.

Apart from the dilemma with the selection of the CF, there are two natural ways of how to construct the pricing formula using the CF. The first one is based on the general (call option) pricing formula

$$\begin{aligned} C(t, T) &= X(t)e^{-q(T-t)}\mathbb{S}[\ln X(T) > \ln K] - Ke^{-r(T-t)}\mathbb{Q}[\ln X(T) > \ln K] \\ &= X(t)e^{-q(T-t)}\Pi_1 - Ke^{-r(T-t)}\Pi_2, \end{aligned} \quad (\text{II.1.16})$$

where the probabilities Π_1, Π_2 are calculated by means of the CF \rightarrow CDF inversion according to (Gil-Pelaez, 1951). The main Gil-Pelaez result is that given an arbitrary CF φ_Y , the CDF $F_Y(y) = \Pr[Y \leq y]$ and the complementary CDF $F_Y^c(y) = \Pr[Y > y]$ equal

$$\begin{aligned} F_Y(y) &= \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \text{Re} \left\{ \frac{e^{-iuy} \varphi_Y(u)}{iu} \right\} du \\ F_Y^c(y) &= \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \text{Re} \left\{ \frac{e^{-iuy} \varphi_Y(u)}{iu} \right\} du. \end{aligned}$$

Clearly, for the illustrated pricing problem (II.1.16), the complementary CDF is the right expression to be used and thus for $j = 1, 2$ we have

$$\Pi_j = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \text{Re} \left\{ \frac{e^{-iu \ln K} \varphi_j(u)}{iu} \right\} du,$$

where

$$\begin{aligned} \varphi_j(u) &= \exp(C_j(u) + D_j(u)v(t) + iu \ln X(t)) \\ D_j(u) &= \frac{b_j - \rho\sigma iu - d_j(u)}{\sigma^2} \left(\frac{1 - e^{-d_j(u)(T-t)}}{1 - c_j(u)e^{-d_j(u)(T-t)}} \right) \\ C_j(u) &= (r - q)iu(T - t) \\ &+ \frac{\kappa\theta}{\sigma^2} \left[(b_j(u) - \rho\sigma iu - d_j(u))(T - t) - 2 \ln \left(\frac{1 - c_j(u)e^{-d_j(u)(T-t)}}{1 - c_j(u)} \right) \right] \\ c_j(u) &= \frac{b_j - \rho\sigma iu - d_j(u)}{b_j - \rho\sigma iu + d_j(u)} \\ d_j(u) &= \sqrt{(\rho\sigma iu - b_j)^2 - \sigma^2(2m_j iu - u^2)}, \end{aligned}$$

given the constants

$$m_1 = \frac{1}{2}, m_2 = -\frac{1}{2}, b_1 = \kappa - \rho\sigma, b_2 = \kappa.$$

Here, we note that φ_1 instead of being defined through the specification above, can alternatively be defined via

$$\varphi_1(u) = \frac{\varphi_2(u - i)}{\varphi_2(-u)},$$

and so definition of the CF φ_2 would only be needed in practice.

The main contribution of (Albrecher et al., 2007) is using $c_j(u)$ rather than⁵² $g_j(u) = 1/c_j(u)$ and simple algebraic adjustments in $D_j(u)$. Albrecher's propositions have already been reflected in the expressions above. The obvious reason why we have two characteristic functions φ_1, φ_2 is that they represent the \mathbb{S} and the \mathbb{Q} distribution of $x(T) = \ln X(T)$, respectively. Hence, if one sticks with the approach of computing both Π_1, Π_2 separately he has to numerically evaluate the Gil-Pelaez integral twice, one evaluation for each characteristic function $\varphi_j, j = 1, 2$. This implies two negative aspects: (i) it can slow down the computation of the price because the numerical integration is time-consuming. And more importantly (ii) the errors resulting from the two numerical integrations will add up.

It is therefore beneficial to combine both integrals into a single integral to end up with the following expression for a call

$$\begin{aligned} C(t, T) &= \frac{1}{2}(X(t)e^{-q(T-t)} - Ke^{-r(T-t)}) \\ &+ \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \left\{ \frac{e^{-iu \ln K}}{iu} \left(X(t)e^{-q(T-t)}\varphi_1(u) - Ke^{-r(T-t)}\varphi_2(u) \right) \right\} du. \end{aligned} \tag{II.1.17}$$

The put price can be then easily recovered through the put-call parity⁵³

$$P(t, T) = C(t, T) + Ke^{-r(T-t)} - X(t)e^{-q(T-t)}.$$

We stick with the combined formula (II.1.17) because this only involves a single integration and the computation also works with a smaller error. Finally, we should discuss the numerical integration itself. We in general recommend using Gaussian quadrature over Newton-Cotes formulas because the Gaussian quadrature can better capture the asymptotic behaviour of the integrand. In particular we recommend using Gauss-Laguerre quadrature which performs very well for the problem posed, since this kind of quadrature is natively designed for integrations over $(0, \infty)$. Possibly, one could also employ Gauss-Legendre quadrature which integrates over a general region (a, b) such that the the integrand is evaluated in neither of the endpoints a, b . In the applications, we apply 32-point Gauss-Laguerre quadrature.

⁵²the form of the CF in terms of $g_j(u)$ frequently appears in the literature.

⁵³notice that the put-call parity, when applied to the expression (II.1.16), implies a nice formula for the put as

$$P(t, T) = Ke^{-r(T-t)}(1 - \Pi_2) - X(t)e^{-q(T-t)}(1 - \Pi_1),$$

II.1.5 The analysis

Having presented several possible firm-value model specifications we will use them in the analysis. The main objective of the analysis is to compare the abilities of the models to explain the credit spreads on the market. This, however, has previously been done by many authors and our aim is *not* to follow them. (Beem, 2010) tries to find linkages between CDS (Dutch companies) produced by the jump-diffusion structural models and macro variables, (Zhang et al., 2009) focuses on a decomposition of the variation in CDS quotes into jump and volatility component of equity returns, (Haworth et al., 2008) examines the effect of the credit contagion on CDS quotes in a structural model framework.

The key difference between our approach and the approaches applied in other studies is that we calibrate the models using selected option quotes rather than just the stock market. This is a fundamentally different approach as we expect the option prices to provide a forward-looking view on the company and not just the point-in-time view provided by the stock market.⁵⁴ The closest analysis to ours is that of (Hull et al., 2004), who apply Geske's model to find connections between stock options and CDS.

We now provide a greater insight into what this particular study is about. We start with a motivation which consists of four points:

- (i) The traditional approach to calibration of the Merton model usually underestimates the CDS spread for the subinvestment grade names.
- (ii) If a more complex setup (relative to Black-Scholes) is to be selected, then there are typically several model parameters to be found but it is very difficult to estimate them using traditional approaches as the only observable (stock market) inputs are the equity E and its volatility σ_E .
- (iii) The price of the equity E is a horizon-free number. However, the credit investors take into account the credit risk in the investment horizon. For them it matters what is the term-structure of the default risk involved in the debt. Therefore the equity itself can not be a good predictor of company's health over various horizons.
- (iv) The option quotes on the firm's equity are considered to be a very good predictor of the company's future states. The expectations embedded in a set of liquid option quotes could give a more accurate picture about the future health of the company than just a single horizon-free price of the equity E .

Summarizing all the points (i)-(iv) we analyze whether various structural models that we fit to option quotes outperform the standard Merton model by matching the CDS spreads.

The next key point is how to connect the option quotes and the CDS spreads through the structural models. We apply the key assumption that an option quoted on equity is in view of the structural models a *compound* option, that is option on option on company's

i.e. in terms of the complementary probabilities.

⁵⁴this clearly could be subject to further discussion as many could argue that equity prices also reflect some future outlook on the company. The option quotes, however, provide more direct outlooks as they also implicitly express *probabilities* of the market scenarios to happen with respect to given some horizon T .

assets because option on the company's assets itself is seen as the equity in the structural models.

Our benchmark is the Merton model calibrated via the standard calibration approach described in the end of the Section II.1.4.1. Then we calibrate *compound option* models based on Merton (this is actually analytically tractable *Geske's* model, see (Geske, 1979)), Heston, Zhou (Sections II.1.4.1–II.1.4.4) which can potentially outperform the standard Merton model in their ability to fit the CDS spreads.⁵⁵

II.1.5.1 Proposed general method to the model calibration

In this section we propose a framework that generalizes the ideas of (Geske, 1979) and in particular of (Hull et al., 2004). The key idea is to calibrate the structural credit models of Merton (Geske), Heston and Zhou to a set of put options. Put options have intentionally been chosen as they are more related to the possible equity decrease than the call options, and as our concern is default due to loss of market value of the assets X , put options are a reasonable choice.

The pivotal idea in the proposed calibration framework is to regard the market option quotes on the company's equity E as compound options on call option on the company's assets X . Let us introduce some notation to formulate this idea more explicitly. Let T_1 be a time of maturity of the quoted option on the equity E with a strike K_1 and let T_2 be the maturity of the company's hypothetical zero-bond of the face value K_2 . We further assume $t = 0 < T_1 < T_2$. Since the equity $E(0)$ in the structural models equals the value of the call on the assets X , we set

$$E(0) = C(0, X(0), T_2, K_2).$$

Using this identity a market-quoted option V (be it call or put) on the equity E , maturing at T_1 with a strike K_1 can be written as

$$V(0, E(0), T_1, K_1) = V(0, C(0, \theta_1, T_2, K_2), T_1, K_1 | \Theta),$$

where we have added a parameter vector $\Theta = (\theta_1 = X(0), \theta_2, \dots)$ related to dynamics of the asset process X under \mathbb{Q} . Let us assume having a set of N option market quotes $V_n^M(T_1^{(n)}, K_1^{(n)})$, $n = 1, \dots, N$ on the company's equity E . Then the parameter set $\hat{\Theta}$ that gives the best match to the option market data is a solution to the following optimization problem⁵⁶

$$\hat{\Theta} = \arg \min_{\Theta} \left[\sum_{n=1}^N |\varepsilon_n(\Theta)| \right], \quad (\text{II.1.18})$$

where

$$\varepsilon_n(\Theta) = \frac{V(0, C(0, \theta_1, T_2, K_2), T_1^{(n)}, K_1^{(n)} | \Theta) - V_n^M(T_1^{(n)}, K_1^{(n)})}{V_n^M(T_1^{(n)}, K_1^{(n)})}.$$

⁵⁵for the purpose of this exercise we omit the Merton model with stochastic r presented in the Section II.1.4.2 as the focus here is on modeling 5Y CDS spreads where the contribution of the interest rates is rather negligible.

⁵⁶we have chosen the total relative error as the minimized function in order to account for the importance of the OTM quotes.

This optimization problem, however, can be only solved when one knows how to calculate the value of the option on call $V(0, C(0, \theta_1, T_2, K_2), T_1^{(n)}, K_1^{(n)} | \Theta)$ and this is indeed a delicate task. In fact the only closed-form solution for V exists in the Black-Scholes setup and was discovered by (Geske, 1979). We have customized his result so that it can be written as a one-liner.

Theorem II.1.4 (Compact formula for Geske’s model)

Let $V(t)$ be the price of a compound option with strike K_1 maturing at T_1 on an option with strike K_2 maturing at $T_2 > T_1$, whose underlying is some asset X with volatility σ , subject to GBM. Then the price of the compound option V can be generally written as

$$\begin{aligned} V(t) &= w_1 w_2 X(t) e^{-q(T_2-t)} M(w_1 w_2 a_1, w_2 b_1 | w_1 \rho) \\ &- w_1 w_2 K_2 e^{-r(T_2-t)} M(w_1 w_2 a_2, w_2 b_2 | w_1 \rho) - w_1 e^{-r(T_1-t)} K_1 N(w_1 w_2 a_2), \end{aligned}$$

where

$$\begin{aligned} a_1 &= \frac{\ln\left(\frac{X(t)}{X^*}\right) + (r - q + \frac{1}{2}\sigma^2)(T_1 - t)}{\sigma\sqrt{T_1 - t}}, & a_2 &= a_1 - \sigma\sqrt{T_1 - t}, \\ b_1 &= \frac{\ln\left(\frac{X(t)}{K_2}\right) + (r - q + \frac{1}{2}\sigma^2)(T_2 - t)}{\sigma\sqrt{T_2 - t}}, & b_2 &= b_1 - \sigma\sqrt{T_2 - t}, \\ \rho &= \sqrt{\frac{T_1 - t}{T_2 - t}}, \end{aligned}$$

and M is a bivariate standard normal CDF. Here X^* is the critical value of X for which the Black-Scholes option price with payoff given by w_2 and strike K_2 and residual maturity $T_2 - T_1$ has value equal to K_1 . The variable w_1 is 1 (-1) if the compound option is call (put) and w_2 is 1 (-1) if the underlying option is call (put).

For approaches *not* based on Black-Scholes, we, however, need a more robust and model-free calibration methodology. The methodology we propose only requires a model to be at least semi-analytically tractable for *vanilla plain* option pricing problems.

The idea to compute V in such cases is to simulate all the necessary processes up to T_1 . Standing at T_1 price the call on assets $C(T_1, X(T_1), T_2, K_2)$ using the semi-analytical formula and compute the value of compound option on $C(\cdot, X(\cdot), T_2, K_2)$. Discounting over all scenarios back to $t = 0$ (from T_1) gives the correct value of the compound V . Any simulation, however, would highly disturb the optimization (II.1.18) because any randomness within the optimization is unwanted. The trick we propose is to simulate the random values upfront and holding fixed these pre-generated random numbers run (II.1.18). This half-Monte-Carlo approach allows for very large class of models especially if we take into account the models that are solvable in terms of the Fourier transform approaches (which give semi-analytical solutions).

An alternative method to the simulations is through PDEs/PIDEs. A compound option in a PDE/PIDE framework satisfies the same PDE/PIDE as the underlying option with the only distinction in the terminal condition. For example, assume a put-on-call in the Zhou model. Then this put-on-call P satisfies

$$\frac{\partial P}{\partial t} + (r - q - \tilde{\lambda}\tilde{\beta})X \frac{\partial P}{\partial X} + \frac{1}{2}X^2\sigma_x^2 \frac{\partial^2 P}{\partial X^2} + \tilde{\lambda} \int (P(t, zX) - P(t, X)) f_J(z) dz - rP = 0,$$

where $t \in [0, T_1]$, f_J is the log-normal density and the T_1 -terminal condition is set to

$$P(T_1, X) = \max[K_1 - C(T_1, X, T_2, K_2), 0].$$

Here, C is a the Zhou's analytical formula for the call on the assets X (that is C represents the equity) and thus the 'hard' computations reduce only to the one dimensional PIDE for P . We, however, stick with the simulation-based approach.

The last part of this section is dedicated to a discussion about specific issues of the model calibration. First of all allow us to comment the selection of options whose quotes V^M are to be matched within (II.1.18). One shall, in general, use the most liquid quotes. We in particular recommend using close-to-ATM options whose maturity is furthermore close to 1 year. These are usually very liquid and carry a decent amount of market expectations about the future of the firm. Secondly, as already briefly outlined above, we recommend to use *put option* quotes. A put option is an instrument betting on the decline of the underlying and thus it is more suitable for analysis of default due to low value of the firm (and thus also its equity) as it primarily carries the expectations of the underlying to decline. Our second remark is towards the dividend rate often appearing in the option models (and embedded in option prices). One should bear in mind here that the (continuous) dividend rate is related to the equity (not the assets). If one wants to include the dividends into the compound option model in the firm-value setup, he has to scale the dividend yield such that the cash dividend yield on the assets X matches that of the equity E . Thus, if the dividend yield of the equity is q_E , then the dividend yield q applied in the asset dynamics shall be approximately $q = q_E E/X$. Since $E/X < 1$, the dividend yield on assets has to be strictly lower than the dividend yield on the equity. For companies with a very high leverage (assets to equity) this means that q converges to 0. In fact, the contribution of q for the dynamics of X is only marginal and thus it is relaxed in the analysis.

II.1.5.2 The data

For the purpose of the analysis, we have collected data of 31 US-based companies for which both the equity put options and 5Y CDS spreads are quoted. All the collected data are as of February 15, 2016. We selected representatives from various classes of credit quality, therefore both the investment and the subinvestment grade companies are included in the dataset. Details about the CDS spread of each firm in the dataset and the firm's leverage are shown in the Figure 2. Summarizing the data then results into corresponding histograms in the Figure 3. It is easy to see that the highest CDS quote was about 3000 bps (company #30) and the most levered company was the company #2 with the leverage of nearly 12. It is interesting to observe from the Figure 2 that only the leverage itself can not be the unique explanatory variable in a credit model. A company #30 has a leverage of 3 but the CDS trades at 3000 bps, while the company #4 and #5 have leverage between 6 and 8 and their CDS quotes are just about 100 bps. These are typical examples of banks which are relatively levered yet their debt might still trade at low spread.

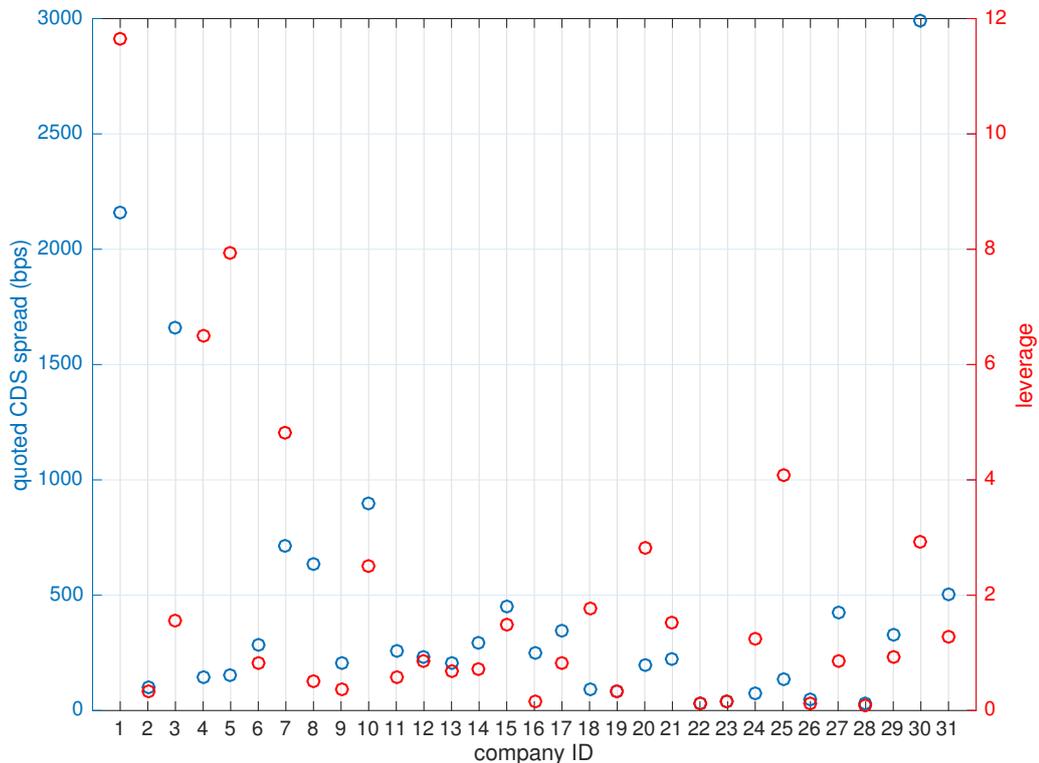


Figure 2: CDS spread and leverage for every company $i = 1, \dots, 31$, in the dataset.

The collected option data cover 487 *put* option quotes with strikes close to ATM, where the majority matures on Jan 20, 2017, i.e. relative to the pricing date, the residual time to maturity is slightly below 1 year. For every company there were at least 6 put quotes which also guarantees sufficient number of quotes for calibration of more complicated models. The provided pricing quotes were for bid, ask and last. We completely dismissed the 'last trade' prices as these are in the case of less liquid options typically older than the bid and the ask quotes. We finally simply averaged the bid and ask quotes into mid quotes, these are henceforth referred to as 'option prices' or 'quotes'.

We also collected the balance sheet data and some additional mixed data. The balance sheet data is the most recent *total* debt of the company, the mixed data is the outstanding number of shares, average debt maturity, share price and its volatility. For convenience we expressed all the data in a per-share basis, i.e. the total debt was divided by the shares out. Also, we did *not* classify the debt into the short-term and long-term debt and thus did not adjusted the short-term debt (as it is customary e.g. in the KMV model). Instead, we directly used the average debt maturity⁵⁷ which we furthermore scaled, capped and floored. Let τ denote the average debt maturity as published by the agency

⁵⁷according to the agency providing the data, this number was calculated as a volume-weighted maturity of the debt.

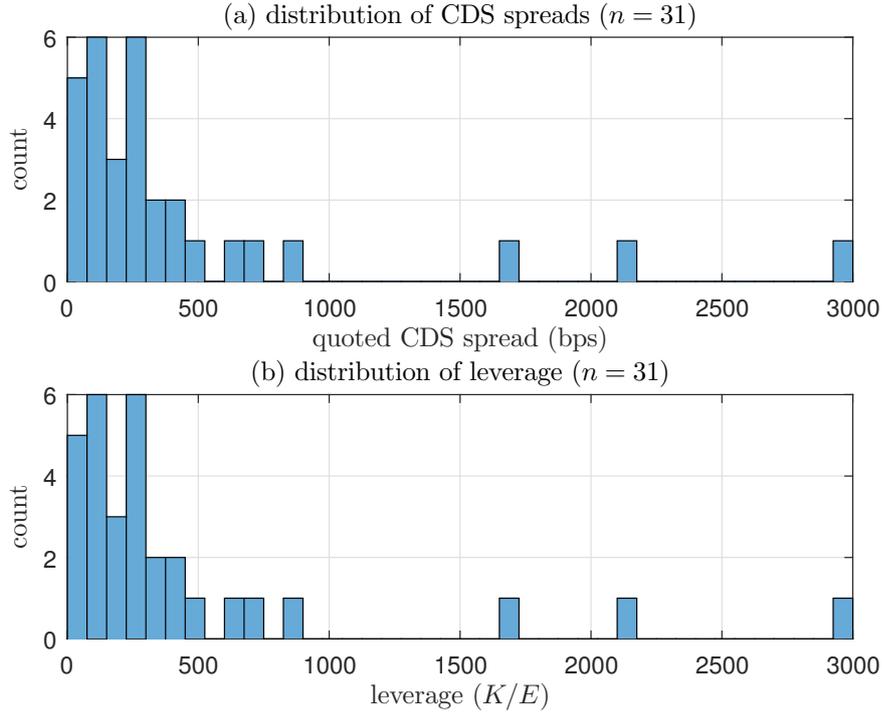


Figure 3: Distribution of companies according to CDS spreads (a) and leverage (b).

and let T_2 denote the hypothetical debt maturity later used in the model. We observed that it is optimal to set $T_2 = 5\mathbf{1}_{\{\alpha\tau \leq 5\}} + \alpha\tau\mathbf{1}_{\{5 < \alpha\tau < 10\}} + 10\mathbf{1}_{\{\alpha\tau \geq 10\}}$, where $\alpha = 1.2$ for banks and $\alpha = 1.5$ for other companies. This guarantees $T_2 \in [5, 10]$. Under this setup, we discovered that the yield spreads populated by the models are neither systematically undershooting nor overshooting the real-world CDS spreads and so the difference model yield spread – CDS spread oscillates around 0.

II.1.5.3 The results

The Merton model (standard calibration). In the Mertonian setup, we applied the classical $(X(0), \sigma_x)$ estimation procedure from the equity data as described in the Section II.1.5.1. (The calibration of the 'Merton' model on the option data is then discussed under the Geske model). The results we have obtained are depicted in the Figure 4. From this picture it is easy to see that the model systematically underestimates the real-world credit spreads.⁵⁸ In the majority of the cases, the model was unable to explain the riskiness of the debt by means of a reasonable implied yield spread for pricing. The model, although not suitable for the pricing, was however surprisingly accurate in classification of companies into 'very risky' and 'not very risky'. Let us define the objective level of *very risky* as having quoted 5Y CDS higher than 500 bps and anything below this level as *not very*

⁵⁸regardless of the choice of the debt's hypothetical maturity T_2 , the spreads implied by the Merton model were always undershooting the real-world CDS quotes.

risky. Let us define the *model very risky* as the case with implied yield spread higher than 200 bps and *model not very risky* anything below 200 bps. Then the calibrated Merton model gave the hit ratio as shown in the Table 1. It successfully classified 23 companies as not very risky out of 24 not very risky companies in the sample, this is 95.8% success rate. In the very risky cases there were 7 very risky cases observed of which the model 6 classified as very risky. This is 85.7% success rate. Therefore the model definitely entails some classification power and so for the practical use the outputs need to be mapped to some real-world cases. This is indeed the idea behind the KMV's implementation of the Merton model with the difference that KMV model does not concern CDS but rather default probabilities.

model/observed	not very risky	very risky
not very risky	23	1
very risky	1	6
total	24	7

Table 1: Successful and unsuccessful hits of the Merton model.

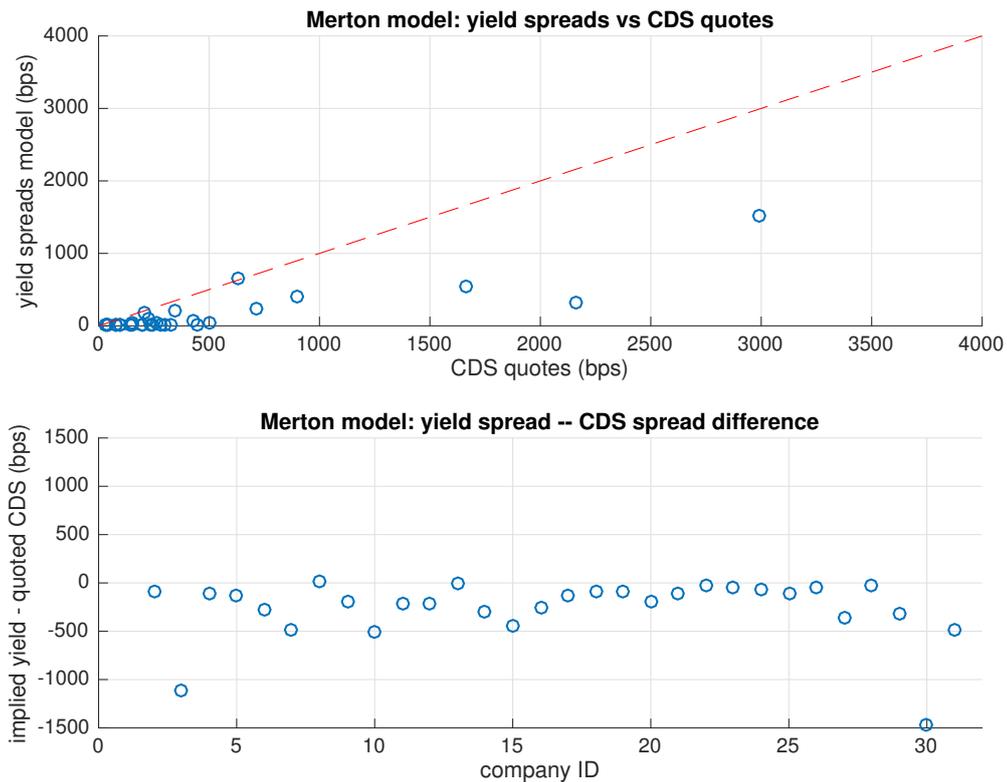


Figure 4: Results in the Merton model. The upper chart shows model yields vs. quoted CDS spreads. The red line is a theoretical 'perfect-fit' line. The lower chart shows differences: model spread – quoted CDS spread for all companies i in the dataset ($i = 1, \dots, 31$).

The Geske model. The Geske model has been applied through the option-on-option

implementation with the good property that there exists an analytical solution for such an option, therefore there was no need for any simulation. This makes the model very fast to be deployed and used. We find that the 'put option on equity = put-on-call on assets' approach provided much more consistent and balanced results for pricing compared to the standard Merton model. The outputs are graphically presented in the Figure 5. We can easily see that the model is able to populate yield spreads that mimic that of the CDS quotes. The generated yields oscillate around the perfect-fit line which is a good sign, as these points are not too far from the line. The only drawback we observe is that although the model is arguably better for pricing than the Merton model, it has no particularly good performance for neither the 'very risky' nor the 'not very risky' names. In particular in the case of names with 5Y CDS lower than 500 bps (24 names), the model spreads were in 18 cases (75% of 24) lower which is to be seen in the chart. Hence, for the not very risky names it seems to underestimate the yield spread.

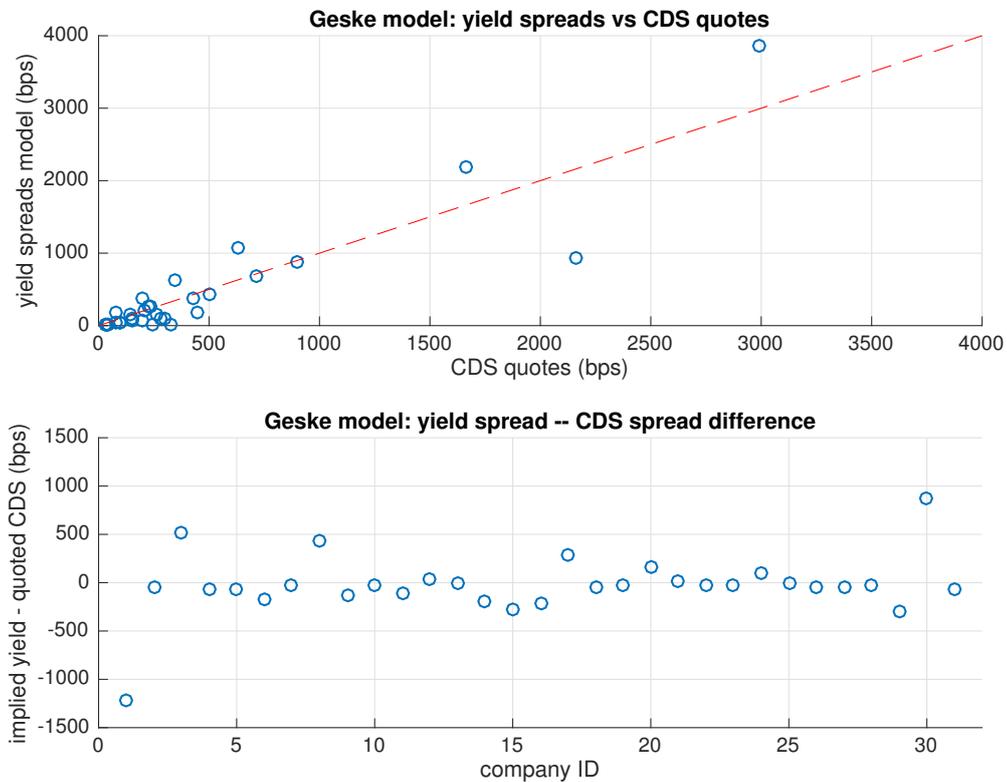


Figure 5: Results in the Geske model. The upper chart shows model yields vs. quoted CDS spreads. The red line is a theoretical 'perfect-fit' line. The lower chart shows differences: model spread - quoted CDS spread for all companies i in the dataset ($i = 1, \dots, 31$).

The Hestonian approach. From the selection of the models we tested, the Hestonian implementation of the option-on-option approach to credit risk was the most technically demanding one. We followed the calibration instructions in Section II.1.5.1 but there was to do extra performance tuning in order to calibrate the model. The technical complication

with the Hestonian implementation of the firm value model according to the selected option-on-option approach is that hundreds of millions evaluations of the formula for call option need to be run.⁵⁹ This, however, involves the numerical integration which makes the process very slow. We measured that with no optimization just 100 000 evaluations of the pricing formula took 3 minutes. This would mean that calibration to one company's data would take several days. We therefore conducted the following optimization steps. Firstly, we vectorized the characteristic function φ_j such that it can be instantaneously evaluated at all the integration points. The second adjustment is to allow the pricing function to instantaneously calculate the prices of one option for various values of the simulated pairs $\{X_j(t), v_j(t)\}, j = 1, \dots, J$ all other inputs (parameters) being equal. All this means to create matrices for $X(t), v(t)$, integration nodes and integration weights.⁶⁰ In total all these matrices are of the dimension $(J \times 32)$. After the option function was able to instantaneously calculate prices for various $X(t), v(t)$, the execution time has decreased dramatically to 1.2% of the original time. This implies that now it would take an hour or two to calibrate the model to one company's data. We additionally employed the GPU⁶¹ to run the big array computations on the computer's graphic card. This has brought another significant (5 \times) speed boost decreasing the total runtime to just 0.2% of the original runtime. A final tuning was that we allowed for parallel computations making the total runtime only 0.13% of the initial runtime. In total, the calibration to one company's data with this setup took always just a few minutes.

After a successful calibration, we ran the model and received the results presented in the Figure 6. The model exhibits undershooting of the low spreads and the populated spreads for 'not very risky' names are worse than in the case of Geske's model. For the 'risky' names the model seems to perform more consistently and suprisingly it was almost able to fit the 3000 bps CDS quote. This is in particular a very good result as we expect the CDS quotes of the junk names to be increasingly difficult to be fit.

⁵⁹if one assumes 30 000 simulations, 10 option quotes for a company i and 2000 iterations of the optimization algorithm, then there are $30\,000 \times 10 \times 2\,000 = 600\,000\,000$ evaluations of the option pricing formula.

⁶⁰notice we use 32-point Gauss-Laguerre quadrature.

⁶¹graphics processing unit (graphic card). GPU is much faster than CPU for some specific tasks.

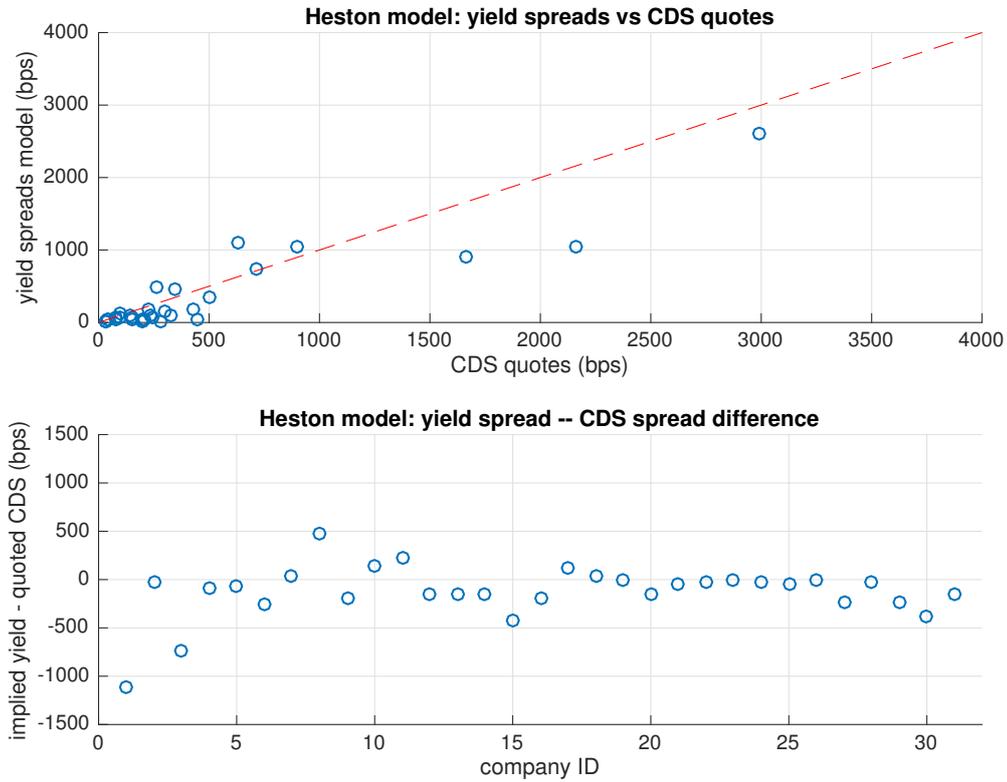


Figure 6: Results in the Heston model. The upper chart shows model yields vs. quoted CDS spreads. The red line is a theoretical 'perfect-fit' line. The lower chart shows differences: model spread – quoted CDS spread for all companies i in the dataset ($i = 1, \dots, 31$).

The Zhou model. Finally, we also performed the option-on-option calibration within the Zhou's jump diffusion model. Although it was not technically as complicated as the Heston case, still a significant optimization in the algorithm was needed to minimize the runtime. The Zhou's model (Figure 7) has overall a relatively good fit to the observed CDS spreads. It is, however, convincingly accurate for the 'not very risky' cases where it excels among all the models in the analysis. For these cases, the points in the upper chart of the Figure 7 are closely dispersed around the perfect-fit line. Excluding the obvious one outlier, the average *absolute* error was only 83 bps, while the average difference (not absolute) from the CDS quotes was only -12 bps. For the risky group, the model performed rather moderately, yet providing still relatively balanced results.



Figure 7: Results in the Zhou model. The upper chart shows model yields vs. quoted CDS spreads. The red line is a theoretical 'perfect-fit' line. The lower chart shows differences: model spread – quoted CDS spread for all companies i in the dataset ($i = 1, \dots, 31$).

II.1.5.4 Conclusion

The objective of the Section II.1.5 was to test all the introduced models on a real-world sample data collected from 31 companies. The key finding is the following. Equity (put) option quotes carry a significant amount of evidence of credit risk of the company. In all the presented models, the approach to calibration of firm-value models through option quotes totally outperformed the outcomes from the benchmark standard Merton model. Therefore, the option-on-option approach seems to be an adequate technique for pricing the credit risk, even though options have virtually nothing to do with the company's credit.

Our second finding is that there is, in general, no universally good model for assessment of the company's credit risk, yet a reasonable combination of various models might lead to very accurate results. The Merton model calibrated through the standard method to only equity data and debt as the inputs, was able to discriminate the 'very risky' and 'not very risky' companies with a high success rate, yet the implied credit spreads were dubious. The other approaches based on option-on-option method were much more successful in matching the CDS spreads. In particular, the Geske's model offers a good tradeoff between the general precision and analytical tractability. The Hestonian implementation failed to match the low spreads but showed a good performance for the high yield credits. Finally,

our implementation of the Zhou’s model exhibits excellent results for the investment grade credits.

An interesting look at the results might be from an investor’s perspective. If we adopt the idea that the models are, in general, correct, then if all models simultaneously show a significantly under/overpriced credit, this might indicate an investment opportunity. One such possibly wrongly priced credit might be the one with the quote 2163 bps (company #1). For this company, all models uniformly indicate a much lower spread (about 1000 bps) than what can be observed in the market. This, however, might be tricky because the quality of the model outcome is subject to the quality of the option quotes. If the option quotes are illiquid or noisy, the credit analysis obviously provides misleading figures. One should therefore also conduct a deep analysis of the data integrity.

Last but not least we shall stress out that if the option quotes are relatively sparse one could, possibly, generate the whole volatility surface by employing some state-of-the art method such as SSVI (Surface Stochastic Volatility–Inspired) pioneered by Jim Gatheral, which allows to fit the shape of the implied volatilities across the (T, K) –space. A surprisingly simple parametric form of the implied volatility (IV) function

$$\sigma_{\text{IV}}(T, K) = \sqrt{\frac{\theta_T}{2T} \left(1 + \rho\varphi(\theta_T)\kappa(K, F_T) + \sqrt{(\varphi(\theta_T)\kappa(K, F_T) + \rho)^2 + (1 - \rho^2)} \right)},$$

is in the most cases able fit to the observed IV quotes and then this function $\sigma_{\text{IV}}(T, K)$ can be used to find IVs for arbitrary combinations of (T, K) . A more general discussion is provided by (Gatheral, 2006) and SSVI in particular is then outlined in (Gatheral and Jacquier, 2014, Section 4).

Part III

Reduced–form models

In this part we will introduce a class of credit risk models referred to as *reduced-form* models. The main difference between the firm-value models and the reduced-form models is that the reduced-form models do not assume any particular form of the default trigger and so the default comes as a total surprise. This clearly is a great advantage over the asset value models in the sense that no stringent and often criticised assumption of firm’s assets being traded is needed. The default probability in the reduced-form models is typically brought in by some ‘intensity’ (or hazard) function or process. In spite of the apparent benefits of the reduced-form models, there are clearly some limitations or disadvantages of these models, relative to the firm-value approaches. The major disadvantage is that the asset value models are based on the empirically validated observation that a company defaults when it is unable to pay its debt. This explicit treatment of default, however, does not exist for the reduced-form models. Also, the treatment of the default correlations is relatively limited for the reduced-form models although there are some extensions one of which we introduce in this text.

This Part III is organized as follows. In the Section III.1 we outline the main theoretical framework for the reduced-form modeling. This is essentially a mix of theory that can be per-partes found in (Jarrow and Turnbull, 1995), (Jarrow et al., 1997), (Lando, 1998), (Bielecki and Rutkowski, 2004), (Jarrow and Protter, 2004) and many other texts yet the theory in this literature is often not complete or maybe focuses too much on the measure-theoretical concepts at the expense of linkages to practical applications. The main building blocks of the Section III.1 are the construction of the default time, flow of credit and market information, hazard function and hazard process.

Based on the theory related to the intensity processes,⁶² found mainly in (Lando, 1998) the Section III.2 we derive a complete general PDE framework for the reduced-form credit models and directly propose various useful PDE problem definitions (and solutions) to many real-world modeling problems and supplement this with two practical calibration methods of the reduced-form models to the market data. We also present a CVA extension and a more elaborate framework where jumps in the default intensity are allowed for, resulting into partial integro-differential equations (PIDEs). The work with PDEs/PIDEs has also led us to a special PDE/PIDE-based method of estimation of parameters in (jump)-diffusion models. This can, potentially, be useful for the estimation of joint dynamics and/or correlation of processes. Although this is not necessarily related to the reduced-form modeling, this approach was initially motivated by the (joint) estimation of default intensities. The method is presented in the Appendix D. Since the most problems we deal with in the Section III.2 require numerical solutions to PDEs, for this purpose we have defined a suitable (finite difference method) numerical schemes. These schemes are completely explained in the Appendix C.

In the Section III.3 we move on to a self-developed binomial tree framework for pric-

⁶²(Lando, 1998) refers to a stochastic intensity process as a *Cox process*.

ing of defaultable securities. The essential concept mainly follows the setup of (Jarrow and Turnbull, 1995) but we present a much more general framework, where the tree is constructed using a two-level lattice. This then allows for pricing of more complicated securities such as barrier options on defaultable bonds. At the end of this section, to validate the results and to demonstrate how the modeling techniques are connected we perform a numerical verification through the results provided in the earlier Section III.2 by combining PDEs and MC.

III.1 Definitions and essential concepts

In the reduced-form framework, we in general work under the following setup. We assume that all the information can be captured by two information flows. The first information flow $\mathcal{H}(t) = \sigma(\mathbf{1}_{\{\tau^* \leq u\}} : 0 \leq u \leq t)$ is referred to as the *credit information flow* generated by the default indicator $H(t) = \mathbf{1}_{\{\tau^* \leq t\}}$. It exclusively holds the information about whether the default has occurred or not. Secondly, there exists a *market information flow* $\mathcal{F}(t) = \sigma(X(u) : 0 \leq u \leq t)$ generated by some possibly vector-valued stochastic process X . Typically, the components of X are the risk-free rate process $r = X_1$ and the intensity process $\gamma = X_2$. We also introduce a *combined information flow* defined via $\mathcal{G}(t) = \mathcal{H}(t) \vee \mathcal{F}(t)$ which holds both the information about the default and the market variables. It will be shown that the market information flow plays a critical role for the probability of default when the default intensity γ is stochastic and no role when the intensity is deterministic (or constant).

Also see the fundamental difference between the informational structure between the firm-value and the reduced-form models. In the firm-value models, we just worked with some general filtration \mathbb{G} generated by the firm-value process and other market processes. Hence, there was no need to introduce the filtration \mathbb{H} because the default indicator was measurable with respect to the information generated by the market processes.⁶³ This setup, however, is not optimal for the reduced-form models, where the market processes say nothing about the default. In turn, it is useful to decompose the general filtration \mathbb{G} into smaller informational subsets, \mathbb{H} and \mathbb{F} which exclusively hold information about the default and the market processes, respectively.

We start with the basic definition of the (cumulative) default intensity as a fundamental component of all reduced-form models.

Definition III.1.1 (Default intensity and cumulative intensity)

The default intensity (or hazard) γ is a non-negative infinitesimal rate of default defined via the limit

$$\gamma(t) = \lim_{\Delta_t \rightarrow 0} \frac{\mathbb{P}[t < \tau^* \leq t + \Delta_t | \tau^* > t]}{\Delta_t}.$$

The cumulative intensity (or hazard) reads

$$\Gamma(t) = \int_0^t \gamma(u) du.$$

⁶³if the asset value is too low, a default is likely to be observed.

The definition of the hazard rate γ can also clearly be rephrased so that the probability of default (assuming no default until time t) is

$$\mathbb{P}[t < \tau^* \leq t + \Delta_t | \tau^* > t, \mathcal{F}(u)] \approx \gamma(t)\Delta_t, u \geq t, \Delta_t \rightarrow 0,$$

where we added the conditioning on $\mathcal{F}(u), u \geq t$ to support the cases with intensity being stochastic.⁶⁴ The cumulative hazard rate simply expresses how much default risk (represented by the default intensity γ) has been accumulated up to the time t . Having defined the intensity γ and the cumulative intensity Γ we can finally give a precise definition to the default time τ^* in the reduced-form models.

Definition III.1.2 (Default time τ^*)

The default time τ^* is a random variable which originates according to

$$\tau^* = \inf \{t \in \mathbb{R}^+ : \Gamma(t) \geq v\},$$

where v is a unit exponential random variable ($v \sim \exp(1)$) independent of Γ . Based on this definition, the default time τ^* is said to be totally inaccessible and is a \mathbb{H} -stopping time.

Let us explain how the default time τ^* originates according to this definition. It says that τ^* is the smallest time t for which the accumulated hazard $\Gamma(t)$ exceeds an independently tossed exponential random variable with mean 1 (i.e. $\lambda = 1$ for the exponential distribution).⁶⁵ In other words, we know nothing about the default time τ^* even if we know $\Gamma(t), 0 \leq t \leq T^*$ unless the value of the independent random variable $v \sim \exp(1)$ has been revealed. This is the total inaccessibility of the default time τ^* . An equivalent specification of the default time τ^* to that of the one postulated in the Definition III.1.2 is also

$$\tau^* = \inf \{t \in \mathbb{R}^+ : e^{-\Gamma(t)} \leq U\}, U \sim [0, 1],$$

which more explicitly describes the default time τ^* as the smallest time t at which the credit does not manage to *survive*. The default time τ^* provided in the Definition III.1.2 on the other hand stresses out the message that the default occurs as soon as enough hazard is accumulated.

In the two sections that follow we investigate the properties of the deterministic and the stochastic intensity. These sections underpin the whole framework of the reduced-form models. In these sections we also assume that the default intensity (regardless of whether deterministic or stochastic) is absolutely continuous, i.e. it has no discontinuities.

III.1.1 Hazard function and cumulative hazard function

In this section we investigate some important probability-related aspects of the hazard (intensity) function γ and the cumulative hazard (intensity) function Γ . These relations

⁶⁴notice this conditioning was not necessary in the Definition III.1.1 because we derived the intensity from probability, i.e. we assumed to know the the probabilities of default and the intensity was obtained from these probabilities.

⁶⁵notice that quick way of generating v through a uniform distribution is $v \sim -\ln(1 - U), U \sim [0, 1]$.

are implicitly or explicitly used in many modeling frameworks where the simplifying assumption that γ is a deterministic function of time is made. In order to allow for the deterministic intensity, it is necessary to assume that the *market* filtration \mathbb{F} is *trivial* such that $\mathcal{G}(t) = \mathcal{F}(t) \vee \mathcal{H}(t) \equiv \mathcal{H}(t)$. Another, more general way of describing the modeling environment can be achieved by assuming that \mathbb{F} is non-trivial but the functions γ, Γ are independent of this filtration. This non-trivial market filtration can, for instance support a stochastic interest-rate (independent of the default time τ^*). As $\mathcal{F}(t)$ is either trivial or provides no information about the default time τ^* we refrain from any conditioning on $\mathcal{F}(t)$ in the formulas that follow. Having outlined the basic technical assumptions underpinning the deterministic behaviour of γ, Γ , we investigate the important distributional properties of the default time τ^* in terms of the deterministic functions of time γ and Γ .

We start by defining the probability of default and the probability of survival. The probability of default is also the CDF of the default time τ^* . We have

$$\begin{aligned}\mathbb{P}[\tau^* \leq t] &= F(t) = 1 - e^{-\Gamma(t)} \\ \mathbb{P}[\tau^* > t] &= G(t) = 1 - F(t) = e^{-\Gamma(t)}.\end{aligned}$$

In many pricing problems we are also interested in the probability of default occurring in a specific time interval $(t, s]$ (without making any further assumptions), rather than in the probability of default *until* some pre-specified time. It is not too hard to conclude that this can be computed as the difference between the probability of default until s less the probability of default until t . This yields

$$\mathbb{P}[t < \tau^* \leq s] = \mathbb{P}[\tau^* \leq s] - \mathbb{P}[\tau^* \leq t] = e^{-\Gamma(t)} - e^{-\Gamma(s)}.$$

Furthermore, we enrich the probabilities for some additional assumptions about the default time τ^* . More specifically, we assume $\tau^* > t$, i.e. we make the assumption that the credit has survived until t . We immediately have

$$\begin{aligned}\mathbb{P}[\tau^* > s | \tau^* > t] &= \frac{\mathbb{P}[\tau^* > s]}{\mathbb{P}[\tau^* > t]} = e^{\Gamma(t) - \Gamma(s)} = e^{-\int_t^s \gamma(u) du} \\ \mathbb{P}[t < \tau^* \leq s | \tau^* > t] &= \frac{\mathbb{P}[t < \tau^* \leq s]}{\mathbb{P}[\tau^* > t]} = 1 - \frac{\mathbb{P}[\tau^* > s]}{\mathbb{P}[\tau^* > t]} \\ &= 1 - e^{\Gamma(t) - \Gamma(s)} = 1 - e^{-\int_t^s \gamma(u) du},\end{aligned}\tag{III.1.1}$$

which express the *forward* survival and default probability, respectively. These two formulas can be, however, generalized to include the complete information about the default occurring or not by the time t which is captured by $\mathcal{H}(t)$. Notice that $\{\tau^* > t\} \in \mathcal{H}(t)$ used above is just a particular event in $\mathcal{H}(t)$. We find that the general $\mathcal{H}(t)$ -conditional

probabilities are

$$\begin{aligned}
\mathbb{P}[\tau^* > s | \mathcal{H}(t)] &= \mathbf{1}_{\{\tau^* > t\}} \frac{\mathbb{P}[\tau^* > s]}{\mathbb{P}[\tau^* > t]} = \mathbf{1}_{\{\tau^* > t\}} e^{\Gamma(t) - \Gamma(s)} = \mathbf{1}_{\{\tau^* > t\}} e^{-\int_t^s \gamma(u) du} \\
\mathbb{P}[t < \tau^* \leq s | \mathcal{H}(t)] &= \mathbf{1}_{\{\tau^* > t\}} \frac{\mathbb{P}[t < \tau^* \leq s]}{\mathbb{P}[\tau^* > t]} = \mathbf{1}_{\{\tau^* > t\}} \left(1 - \frac{\mathbb{P}[\tau^* > s]}{\mathbb{P}[\tau^* > t]} \right) \\
&= \mathbf{1}_{\{\tau^* > t\}} \left(1 - e^{\Gamma(t) - \Gamma(s)} \right) = \mathbf{1}_{\{\tau^* > t\}} \left(1 - e^{-\int_t^s \gamma(u) du} \right). \quad (\text{III.1.2})
\end{aligned}$$

Notice that the probabilities in (III.1.1) are equal to the conditional probabilities in (III.1.2) evaluated on the set $\{\tau^* > t\}$.

III.1.2 Hazard process and cumulative hazard process

This section is the exact copy of the Section III.1.1 with the only difference that γ and Γ are now *stochastic processes adapted to the market filtration* \mathbb{F} . Under this setup, probabilities conditional on $\mathcal{G}(t) = \mathcal{F}(t) \vee \mathcal{H}(t)$ express the probabilities of default/survival knowing γ and Γ up to the time t and also knowing whether the default has occurred or not by time t . In fact the difference between the expression for the probabilities in terms of the deterministic hazard and the stochastic hazard is that we replace the expressions involving Γ by the 'same expression' under the expectation operator \mathbb{E} . We define the CDF of the default time and the survival probability through the expressions

$$\begin{aligned}
\mathbb{P}[\tau^* \leq t | \mathcal{F}(t)] &= F(t) = \mathbb{E} \left[1 - e^{-\Gamma(t)} \middle| \mathcal{F}(t) \right] = 1 - e^{-\Gamma(t)} \\
\mathbb{P}[\tau^* > t | \mathcal{F}(t)] &= G(t) = 1 - F(t) = \left[e^{-\Gamma(t)} \middle| \mathcal{F}(t) \right] = e^{-\Gamma(t)}, \quad (\text{III.1.3})
\end{aligned}$$

where we assume the market information up to the time t is known (which is captured by $\mathcal{F}(t)$). This can, however, be generalized for the cases when *less market information* than up to the time t is known. In that case we could not replace the expectations $\mathbb{E}[\cdot]$ in (III.1.3) by what is inside of those expectations and the expectation operator would persist. An example would be e.g. $\mathbb{P}[\tau^* > t | \mathcal{F}(0)] = \mathbb{E} \left[1 - e^{-\Gamma(t)} \middle| \mathcal{F}(0) \right]$ where nothing else can be done for the simplification because $\mathcal{F}(0)$ tells nothing about $\Gamma(t)$ and so this remains under the expectation operator. For the unconditional probability of default in $(t, s]$ we have

$$\mathbb{P}[t < \tau^* \leq s | \mathcal{F}(u)] = \mathbb{P}[\tau^* \leq s | \mathcal{F}(u)] - \mathbb{P}[\tau^* \leq t | \mathcal{F}(u)] = \mathbb{E} \left[e^{-\Gamma(t)} - e^{-\Gamma(s)} \middle| \mathcal{F}(u) \right],$$

where we assume that the market information up to some general time u is known.⁶⁶ We also have

$$\begin{aligned}
\mathbb{P}[\tau^* > s | \tau^* > t, \mathcal{F}(u)] &= \mathbb{E} \left[e^{\Gamma(t) - \Gamma(s)} \middle| \mathcal{F}(u) \right] = \mathbb{E} \left[e^{-\int_t^s \gamma(u) du} \middle| \mathcal{F}(u) \right] \\
\mathbb{P}[t < \tau^* \leq s | \tau^* > t, \mathcal{F}(u)] &= \mathbb{E} \left[1 - e^{\Gamma(t) - \Gamma(s)} \middle| \mathcal{F}(u) \right] = \mathbb{E} \left[1 - e^{-\int_t^s \gamma(u) du} \middle| \mathcal{F}(u) \right].
\end{aligned}$$

⁶⁶if $u \geq t$ then $e^{-\Gamma(t)}$ can be taken out of the expectation and if furthermore $u \geq s$ then also $e^{-\Gamma(s)}$ can be taken out of the expectation.

These formulas define the *forward* survival and default probability given the market information up to some general time u . Because these expressions are more difficult to be generalized we generalize them in a special way which assumes *all* information is known up to the time t and we obtain

$$\begin{aligned}\mathbb{P}[\tau^* > s | \mathcal{G}(t)] &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{P}[\tau^* > s | \tau^* > t, \mathcal{F}(t)] \\ &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E} \left[e^{\Gamma(t) - \Gamma(s)} \middle| \mathcal{F}(t) \right] = \mathbf{1}_{\{\tau^* > t\}} \mathbb{E} \left[e^{-\int_t^s \gamma(u) du} \middle| \mathcal{F}(t) \right] \\ \mathbb{P}[t < \tau^* \leq s | \mathcal{G}(t)] &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{P}[t < \tau^* \leq s | \tau^* > t, \mathcal{F}(t)] \\ &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E} \left[1 - e^{\Gamma(t) - \Gamma(s)} \middle| \mathcal{F}(t) \right] = \mathbf{1}_{\{\tau^* > t\}} \mathbb{E} \left[1 - e^{-\int_t^s \gamma(u) du} \middle| \mathcal{F}(t) \right].\end{aligned}$$

In an even more general case we could also assume knowing the credit information up to the time t and the market information up to some general time u . This would involve replacing $\mathcal{G}(t)$ by $\mathcal{H}(t) \vee \mathcal{F}(u)$ and $\mathcal{F}(t)$ by $\mathcal{F}(u)$.

To complete this section, we can not forget to mention the default time *density* assuming $\tau^* > t$ which proves to be very useful in many cases because sometimes it is more convenient to describe τ^* in terms of PDF rather than in terms of CDF. This density reads

$$\frac{\partial}{\partial s} \mathbb{P}[t < \tau^* \leq s | \tau^* > t, \mathcal{F}(u)] = \gamma(s) e^{-\int_t^s \gamma(u) du}, \quad u \geq s,$$

and in order to be able to define it, it is assumed that the *market* information are known *at least* up to the time s . If the hazard rate γ is not stochastic but a deterministic function of time then the same result holds (the conditioning on $\mathcal{F}(u)$ is redundant and can be dropped).

III.2 The fundamental pricing equation

The main motivation for this section is the need for a fairly general pricing equation (PDE) for claims subject to credit risk driven by some, preferably stochastic intensity which would be the reduced-form framework counterpart of the firm-value PDE (II.1.8). This is a delicate task because in the intensity-based credit models typically a sudden jump in the default indicator H induces default. On the other hand, interest rates are usually assumed to follow some continuous stochastic process and thus, intuitively, it seems to be a rather complicated task to connect the default related to a jump with a continuous risk-free rate process. We show that even for such a class of problems there exists a unique PDE formulation.

Let us assume two (possibly correlated) stochastic state variables $x = (x_1, x_2)$, where x_1 represents the risk-free rate r and $x_2 = \gamma$ is the default intensity. This is a sufficient setup for the most claims but an extension to more factors is possible and trivial. Also assume a claim V maturing at T , which could be *either* (i) subject to the credit risk of its issuer or (ii) is written on credit risk. At the maturity T the claim *promises* to pay $g(x_1, x_2)$, where g is a deterministic function of x_1, x_2 . We furthermore assume that

the claim provides its holder with a continuous dividend stream at rate $c(t, x_1, x_2)$ until $\min[\tau^* - , T]$. If there is a default prior to or at maturity T , then the literature usually assumes these two recovery claims to be received by the claimholder:⁶⁷

- a recovery claim $\delta(T, x_1, x_2)$ received at time $t = T$, or
- a recovery claim $Z(\tau^*, x_1, x_2)$ received at $t = \tau^*$. $Z(\tau^*) = 0$ for $\tau^* > T$.

The recovery policy δ is referred to as the Fractional Recovery of Treasury Value (FRTV). We believe that the origins of FRTV are can be found in (Jarrow and Turnbull, 1995). Apart from FRTV, which is rather a convenient mathematical construct,⁶⁸ two other recovery policies are widely acknowledged by the financial literature: (i) the Fractional Recovery of Market Value (FRMV), pioneered by (Duffie and Singleton, 1999), is a policy where a recovery claim in amount of recovery fraction $\phi(\tau^*)$ times the *pre-default MtM* value of the underlying is settled at τ^* , and (ii) the Fractional Recovery of Par Value (FRPV)⁶⁹ where a recovery fraction $\phi(\tau^*)$ of the *notional* is settled at the default time τ^* . To place the policies FRMV and FRPV into a real-world context the FRMV is mostly used in the field of credit risk related to derivative contracts (see for example (Schönbucher, 2003)) while the FRPV is typically used in the bond market in connection with a bond's face value.

The main difference between δ and Z is the fact that the settlement of the recovery payoff is postponed until T in the case of δ but occurs exactly at τ^* for Z . We can therefore see that receiving $\delta(T, x_1, x_2)$ at T is economically equivalent to receiving $B(\tau^*, T)\mathbb{E}^{\mathbb{Q}}[\delta(T, x_1, x_2)|\mathcal{F}(\tau^*)]$ at time τ^* . Since the recovery payoff at time τ^* is to be captured by $Z(\tau^*, x_1, x_2)$ we might consider δ to be redundant⁷⁰ and for this reason $\delta(T, x_1, x_2)$ will not be included in what follows. The recovery policy Z can obviously support any of the introduced recovery policies (FRTV, FRMV or FRPV). We will, however, occasionally use δ as some fixed fraction of the defaulted exposure in such cases Z will be a function of δ .

Following (Lando, 1998) assume that every contingent claim V can be represented in terms of three central components V_g, V_c, V_z . The basic term V_g captures the promised T -terminal payment $g(x_1, x_2)$. V_c models the value of the cash-flow (coupon) payments $c(t, x_1, x_2)$. Finally, the purpose of V_z is to incorporate the recovery payment $Z(\tau^*, x_1, x_2)$

⁶⁷some recovery policies have already been outlined in the Section II.1.1.

⁶⁸it simplifies many otherwise difficult problems as the 'timing' of default τ^* plays no role as the recovery payoff δ always settles at the maturity T with probability 1.

⁶⁹some authors, for example (Duffie and Singleton, 2012), call this policy Fractional Recovery of Face Value (FRFV). Although this shorthand is maybe more appropriate because of the absence of 'PV' in the shorthand (which could indicate it has something to do with present value) FRPV seems to be a more common name for this policy.

⁷⁰this means that the δ is a special case of the recovery Z .

upon default. We define V_g as

$$\begin{aligned} V_g(t, T) &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T x_1(u) du \right) g(x_1(T), x_2(T)) \mathbf{1}_{\{\tau^* > T\}} \middle| \mathcal{F}(t) \right] \\ &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T (x_1(u) + x_2(u)) du \right) g(x_1(T), x_2(T)) \middle| \mathcal{F}(t) \right]. \end{aligned}$$

The cash-flow part V_c capturing the continuously compounded coupon rate $c(t, x_1, x_2)$ is then represented by

$$\begin{aligned} V_c(t, T) &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\int_t^{\tau^*} c(u, x_1, x_2) \exp \left(- \int_t^u x_1(s) ds \right) du \middle| \mathcal{F}(t) \right] \\ &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T \mathbf{1}_{\{\tau^* > u\}} c(u, x_1, x_2) \exp \left(- \int_t^u x_1(s) ds \right) du \middle| \mathcal{F}(t) \right] \\ &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T c(u, x_1, x_2) \exp \left(- \int_t^u (x_1(s) + x_2(s)) ds \right) du \middle| \mathcal{F}(t) \right]. \end{aligned}$$

Finally, the recovery part V_z reads,

$$\begin{aligned} V_z(t, T) &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{\tau^*} x_1(s) ds \right) Z(\tau^*, x_1(\tau^*), x_2(\tau^*)) \middle| \mathcal{F}(t) \right] \\ &= \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\int_t^T x_2(u) Z(u, x_1(u), x_2(u)) \exp \left(- \int_t^u (x_1(s) + x_2(s)) ds \right) du \middle| \mathcal{F}(t) \right]. \end{aligned}$$

Setting $V = V_g + V_c + V_z$ gives a combined expression

$$V(t, T) = \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\eta(t, T) g(x_1(T), x_2(T)) + \int_t^T \eta(t, u) h(u, x_1(u), x_2(u)) du \middle| \mathcal{F}(t) \right], \quad (\text{III.2.1})$$

where

$$\begin{aligned} \eta(t, T) &= \exp \left(- \int_t^T (x_1(u) + x_2(u)) du \right) \\ h(u, x_1, x_2) &= x_2 Z(u, x_1, x_2) + c(u, x_1, x_2). \end{aligned}$$

Assuming $x = (x_1, x_2)$ is Itô's process, the pricing problem (III.2.1) on the set $\{\tau^* > t\}$ by comparison with the Theorem 1.1.5 has a unique Feynman-Kac representation which directly leads to the main statement of this section.

Theorem III.2.1 (The fundamental pricing equation for intensity models)

Assume that the stochastic risk-free rate $x_1 = r$ process and some stochastic intensity process $x_2 = \gamma$ follow under \mathbb{Q} the Itô's processes

$$\begin{aligned} dx_1(t) &= \mu_1(t, x_1, x_2)dt + \sigma_1(t, x_1, x_2)dW_{x_1}^{\mathbb{Q}}(t) \\ dx_2(t) &= \mu_2(t, x_1, x_2)dt + \sigma_2(t, x_1, x_2)dW_{x_2}^{\mathbb{Q}}(t) \\ dW_{x_1}^{\mathbb{Q}}(t)dW_{x_2}^{\mathbb{Q}}(t) &= \rho(t, x_1, x_2)dt. \end{aligned}$$

Then the value of the contingent claim $V(t, T) = V(t, T, x_1, x_2)$ which is subject to default risk is on the set $\{\tau^* > t\}$ a solution the PDE

$$\frac{\partial V(t, x_1, x_2)}{\partial t} + \mathcal{A}V(t, x_1, x_2) + x_2 Z(t, x_1, x_2) + c(t, x_1, x_2) - (x_1 + x_2)V(t, x_1, x_2) = 0, \quad (\text{III.2.2})$$

where

$$\begin{aligned} \mathcal{A} &= \mu_1(t, x_1, x_2) \frac{\partial}{\partial x_1} + \mu_2(t, x_1, x_2) \frac{\partial}{\partial x_2} + \frac{1}{2} \sigma_1^2(t, x_1, x_2) \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} \sigma_2^2(t, x_1, x_2) \frac{\partial^2}{\partial x_2^2} \\ &+ \sigma_1(t, x_1, x_2) \sigma_2(t, x_1, x_2) \rho(t, x_1, x_2) \frac{\partial^2}{\partial x_1 \partial x_2}, \end{aligned}$$

and where furthermore $g(x_1, x_2)$ is a promised claim's payoff at time T , Z is a recovery function paying $Z(\tau^*, x_1, x_2)$ at the default time $\tau^* \leq T$, and $c(t, x_1, x_2)$, $t \in [0, \min[\tau^*-, T]]$ is a continuous coupon rate paid to the claimholder until default τ^* or maturity T whichever occurs first.

Remark. In the reduced-form framework we, in general, recommend to use the whole PDE as a boundary condition. This is because neither Dirichlet nor Neumann boundary conditions can be easily defined in the reduced-form setup.

The linear 2D parabolic convection-diffusion PDE⁷¹ (III.2.2) which we introduced above offers a great degree of flexibility. We show that many real-world pricing problems can conveniently be formulated this way and we also present modified versions of this PDE such that they are numerically optimal. We support this with applications and also show that even a calibration of parameters is relatively straightforward, not requiring any uncommon assumptions. Since we can price various instruments *in the same framework*, the

⁷¹as a 'parabolic' PDE, without any rigour, we regard any second-order PDE lacking (both) $\frac{\partial^2}{\partial x \partial t}$ and $\frac{\partial^2}{\partial t^2}$ terms (this is indeed the case of the PDE (III.2.2) and in general any PDE generated by the Feynman-Kac Theorem I.1.5) but in general there are also PDEs that classify as parabolic even if both $\frac{\partial^2}{\partial x \partial t}$ and $\frac{\partial^2}{\partial t^2}$ appear (some other conditions need to be in effect, though. For more details see (Hoffman and Frankel, 2001, p. 505)). Also (III.2.2) is a 'convection-diffusion' equation (in both dimensions) because $\frac{\partial}{\partial x_i}$ (convection) and $\frac{\partial^2}{\partial x_i^2}$ (diffusion) terms are present ($i = 1, 2$). The literature is, however, not fully harmonized in the definition of the (non)linearity. (Hirsa and Neftci, 2013, Section 12.5) or (Cryer, 2014, p. 362) argues that a PDE classifies as linear if V does not appear among any of the coefficients of the partials (but x_1 or/and x_2 can). A more complete definition is then provided e.g. by (Ugail, 2011, p. 24) who regards PDE as linear when the unknown function V and its derivatives are to the power one everywhere *and furthermore* there are no products between the partials and the unknown function V . This is obviously a stricter statement than that in (Hirsa and Neftci, 2013). We conclude the pricing PDE (III.2.2) is linear, regardless of which of the definitions above is used.

parameters can be extracted from a portfolio of commonly traded (not necessarily the same) products subject to credit risk (typically bonds, CDS).

The PDE has two main branches of applications. It can either be used to calculate the counterparty credit risk in a contingent claim, or it can be used to calculate the value⁷² of credit-risk-related product, such as CDS. Bonds, in particular, belong to both these branches.

Extracting the counterparty credit risk from a transaction has become a significantly challenging discipline after the 2007-2008 crisis.⁷³ In practice, this counterparty credit risk component of a derivative transaction is referred to as Credit Valuation Adjustment (CVA) and this quantity is deduced from a hypothetical risk-free value to obtain the credit-risk adjusted value of the deal. An excellent, comprehensive explanation of this topic is provided by (Gregory, 2012). We present a PDE which naturally allows also to price own credit risk (Debit Valuation Adjustment or DVA, in short) in a derivative transaction, this can be simply done by 'switching' the counterparties of the contract. Surprisingly, this CVA–PDE can be obtained by 're-iterating' some of the results presented earlier. A secondary challenge in the counterparty credit risk modeling is to allow for the wrong/good-way risk to be accounted for. This also can be effectively managed in the PDE framework and more importantly without any further requirements or adjustments. It suffices to allow the process of the intensity γ to be correlated with the process of the risk-free rate r and this way we naturally bring the co-movement or counter-movement of both processes r, γ into the model. This, however, also has its mathematical pitfalls and limitations and we discuss this effect of correlation further in the Section III.2.2.1.

A reader interested in the CVA and DVA discussion is referred to the numerous literature specialized on the CVA, DVA⁷⁴ which has emerged in the recent past. For overview, see for example (Kenyon and Stamm, 2012), (Brigo et al., 2013). (Baran, 2016) discusses the post-crisis pricing approaches, and for deeper focus on pricing of products we refer the reader to (Lichters et al., 2015).

Let us now finally discuss some issues related to the intensity process γ . A very a natural choice of the intensity process is CIR⁷⁵ (see (Cox et al., 1985) for the original article⁷⁶) but as it might cause some unwanted issues in a practical implementation, we dedicate to this process a special treatment. In the later sections (see e.g. the Section III.2.1.3 or the Section III.2.1.4) we will also work with Black-Karasinski which does not suffer from these issues.

The numerical solution to the PDE (III.2.2) can be tricky for two main reasons. First, if $\rho \neq 0$ the PDE can not be solved by standard methods as the mixed derivative term $\frac{\partial^2}{\partial x_1 \partial x_2}$ does not allow for separation of differential operators. Secondly, for intensity close to 0

⁷²or price, by induction.

⁷³a particular focus on the 'post-crisis' valuation of derivatives is for example covered in (Baran, 2016).

⁷⁴the difference of both (non-negative) quantities CVA – DVA is usually referred to as the bilateral CVA (BCVA). The value of the derivative V adjusted for BCVA is then calculated as $V = V_{\text{rf}} - \text{BCVA} = V_{\text{rf}} - \text{CVA} + \text{DVA}$.

⁷⁵CIR ensures that the modeled process stays always positive, given the Feller condition $2\kappa\theta > \sigma^2$ holds.

⁷⁶the model was initially designed to model the evolution of interest rates.

at the inception we might get incorrect results because the lowest value on the 'intensity grid' is close to 0 too. We tackle both problems by introducing the tempered CIR process.

Lemma III.2.1 (The tempered CIR process)

If x follows the CIR process

$$dx(t) = \kappa(\theta - x(t))dt + \sigma\sqrt{x(t)}dW(t),$$

then the tempered CIR process $\tilde{x}(t) = \sqrt{x(t)}$ follows

$$d\tilde{x}(t) = \frac{1}{2} \left(\frac{1}{\tilde{x}(t)} \left(\kappa\theta - \frac{1}{4}\sigma^2 \right) - \kappa\tilde{x}(t) \right) dt + \frac{1}{2}\sigma dW(t). \quad (\text{III.2.3})$$

The main benefit of this alternative CIR representation is that the volatility term is now a constant ($\frac{1}{2}\sigma$), which is a desired feature for the PDE orthogonalization and the secondary benefit is that for low values of the intensity, its square root is not so close to 0 ($\sqrt{\gamma} > \gamma$ for $\gamma < 1$) which allows for a more accurate treatment of low-intensity models without the need to adjust the grid or grid spacing. We now restate the main pricing PDE (Theorem III.2.1) in terms of the tempered CIR (III.2.3).

Theorem III.2.2 (Fundamental PDE subject to tempered CIR intensity)

Assume $x_1 = r$ and $x_2 = \sqrt{\gamma}$ defined under \mathbb{Q} as

$$\begin{aligned} dx_1(t) &= \mu_1(t, x_1, x_2)dt + \sigma_1(t, x_1, x_2)dW_{x_1}^{\mathbb{Q}}(t) \\ dx_2(t) &= \mu_2(t, x_1, x_2)dt + \sigma_2(t, x_1, x_2)dW_{x_2}^{\mathbb{Q}}(t) \\ dW_{x_1}^{\mathbb{Q}}(t)dW_{x_2}^{\mathbb{Q}}(t) &= \rho(t, x_1, x_2)dt, \end{aligned}$$

and the pricing problem (on the set $\{\tau^* > t\}$)

$$\tilde{V}(t, T) = \mathbb{E}^{\mathbb{Q}} \left[\tilde{\eta}(t, T)\tilde{g}(x_1(T), x_2(T)) + \int_t^T \tilde{\eta}(t, u)\tilde{h}(u, x_1(u), x_2(u))du \middle| \mathcal{F}(t) \right],$$

where

$$\begin{aligned} \tilde{g}(x_1(T), x_2(T)) &= g(x_1(T), (x_2(T))^2) \\ \tilde{\eta}(t, T) &= \exp \left(- \int_t^T (x_1(u) + (x_2(u))^2)du \right) \\ \tilde{h}(u, x_1, x_2) &= (x_2(u))^2 Z(u, x_1, (x_2)^2) + c(u, x_1, (x_2)^2). \end{aligned}$$

Then $\tilde{V}(t, T)$ solves

$$\begin{aligned} &\frac{\partial \tilde{V}(t, x_1, x_2)}{\partial t} + \mathcal{A}\tilde{V}(t, x_1, x_2) + (x_2)^2 Z(t, x_1, (x_2)^2) + c(t, x_1, (x_2)^2) \\ &- (x_1 + (x_2)^2)\tilde{V}(t, x_1, x_2) = 0, \end{aligned} \quad (\text{III.2.4})$$

where

$$\begin{aligned} \mathcal{A} &= \mu_1(t, x_1, x_2)\frac{\partial}{\partial x_1} + \mu_2(t, x_1, x_2)\frac{\partial}{\partial x_2} + \frac{1}{2}\sigma_1^2(t, x_1, x_2)\frac{\partial^2}{\partial x_1^2} + \frac{1}{2}\sigma_2^2(t, x_1, x_2)\frac{\partial^2}{\partial x_2^2} \\ &+ \sigma_1(t, x_1, x_2)\sigma_2(t, x_1, x_2)\rho(t, x_1, x_2)\frac{\partial^2}{\partial x_1 \partial x_2}. \end{aligned}$$

This theorem provides a PDE somewhat more applicable than the general PDE (III.2.2). Although it incorporates the tempered CIR, it still contains the unwanted mixed derivative term $\frac{\partial^2}{\partial x_1 \partial x_2}$. We conduct one more transformation in order to eliminate the mixed term. This is referred to as the *orthogonal* version of the PDE (III.2.6).

Theorem III.2.3 (PDE subject to tempered CIR intensity (ortho))

Assume $x_1 = r$ and $x_2 = \sqrt{\gamma}$ such that under \mathbb{Q} these processes follow

$$\begin{aligned} dx_1(t) &= \mu_1(t, x_1, x_2)dt + \sigma_1(t, x_1, x_2)dW_{x_1}^{\mathbb{Q}}(t) \\ dx_2(t) &= \mu_2(t, x_1, x_2)dt + \sigma_2(t, x_1, x_2)dW_{x_2}^{\mathbb{Q}}(t) \\ dW_{x_1}^{\mathbb{Q}}(t)dW_{x_2}^{\mathbb{Q}}(t) &= \rho(t)dt, \end{aligned}$$

and that the orthogonalization condition

$$\frac{\sigma_2(t, x_1, x_2)}{\sigma_1(t, x_1, x_2)} = \zeta(t),$$

is in effect. Set

$$\begin{aligned} a(t) &= -\rho(t) \frac{\sigma_2(t, x_1, x_2)}{\sigma_1(t, x_1, x_2)} = -\rho(t)\zeta(t) \\ y_1 &= x_1 \\ y_2 &= a(t)x_1 + x_2 = a(t)y_1 + x_2. \end{aligned}$$

Then $v(t, T, y_1, y_2) = \tilde{V}(t, T, x_1, x_2)$ (\tilde{V} being stated exactly as in the Theorem III.2.2) is a solution to the boundary value problem

$$\begin{aligned} &\frac{\partial v}{\partial t} + \mu_1^y(t, y_1, y_2) \frac{\partial v}{\partial y_1} + \mu_2^y(t, y_1, y_2) \frac{\partial v}{\partial y_2} \\ &+ \frac{1}{2}(\sigma_1^y(t, y_1, y_2))^2 \frac{\partial^2 v}{\partial y_1^2} + \frac{1}{2}(\sigma_2^y(t, y_1, y_2))^2 \frac{\partial^2 v}{\partial y_2^2} \\ &+ (y_2 - a(t)y_1)^2 Z_y(t, y_1, y_2) + c_y(t, y_1, y_2) - \left(y_1 + (y_2 - a(t)y_1)^2 \right) v = 0, \end{aligned}$$

with the terminal condition

$$\tilde{V}(T, T, x_1, x_2) = g(x_1, (x_2)^2) = v(T, T, y_1, y_2) = g(y_1, (y_2 - a(T)y_1)^2).$$

The PDE is subject to the following components:

$$\begin{aligned} Z_y(t, y_1, y_2) &= Z(t, x_1, (x_2)^2) = Z(t, y_1, (y_2 - a(t)y_1)^2) \\ c_y(t, y_1, y_2) &= c(t, x_1, (x_2)^2) = c(t, y_1, (y_2 - a(t)y_1)^2) \\ \mu_1^y(t, y_1, y_2) &= \mu_1(t, x_1, x_2) = \mu_1(t, y_1, y_2 - a(t)y_1) \\ \mu_2^y(t, y_1, y_2) &= \frac{\partial a(t)}{\partial t} y_1 + a(t)\mu_1^y(t, y_1, y_2) + \mu_2(t, y_1, y_2 - a(t)y_1) \\ \sigma_1^y(t, y_1, y_2) &= \sigma_1(t, y_1, y_2 - a(t)y_1) \\ \sigma_2^y(t, y_1, y_2) &= \sigma_2(t, y_1, y_2 - a(t)y_1) \sqrt{1 - \rho^2(t)}. \end{aligned}$$

Equations of this kind are suitable for the ADI methods because the mixed derivative is

no longer present in the orthogonal PDE. In (Kolman, 2015b) we discuss the main idea behind the orthogonalization and also how these orthogonal PDEs can be efficiently solved.

To this end, particularly in the calibration where one estimates the parameters $\Theta_2 = (\gamma, \kappa_2, \theta_2, \sigma_2)$ low values of γ could cause some stability issues and make the calibration fail completely. To make these complications clearer let us assume we have some optimization algorithm to iterate over various values of these parameters to find the optimal parameter set by matching observed market prices through the PDE. Then such an algorithm typically applies some low and high values of these parameters to find sensitivities. This could easily result into an error if in such an iteration it is applied e.g. $\gamma = 0.002$ (this is not an uncommon value of the intensity, see later e.g. the Table 4 or the Table 5, where we show the parameters calibrated to the observed CDS quotes) even when the tempered CIR model is used because the value and unfortunately also its square root are close to 0 where the grid starts.⁷⁷ It would therefore be desirable to make the intensity grid very dense close to 0. This can effectively be done in terms of log CIR which we introduce in the following lemma.

Lemma III.2.2 (The log CIR process)

If x follows the CIR process

$$dx(t) = \kappa(\theta - x(t))dt + \sigma\sqrt{x(t)}dW(t),$$

then the log CIR process $\tilde{x}(t) = \ln x(t)$ follows

$$d\tilde{x}(t) = e^{-\tilde{x}(t)} \left(\kappa(\theta - e^{\tilde{x}(t)}) - \frac{1}{2}\sigma^2 \right) dt + \sigma\sqrt{e^{-\tilde{x}(t)}}dW(t). \quad (\text{III.2.5})$$

Now return back to the problem we outlined above with the low-intensity setup. Since $\gamma = 0.002$, then $\ln \gamma = -6.2146$ but we can easily always find many lower values of $\ln \gamma$ such that γ is closer to 0 than 0.002. This 'infinite density of points' about $\gamma = 0+$ makes the estimation of parameters much easier. Having introduced the log CIR, we can now redefine the PDE (III.2.2) in terms of $x_2 = \ln \gamma$, i.e. in terms of some general (not necessarily CIR) log intensity process (this is also the case of Black-Karasinski, for example).

Theorem III.2.4 (Fundamental PDE subject to log intensity)

Assume $x_1 = r$ and $x_2 = \ln \gamma$ defined under \mathbb{Q} as

$$\begin{aligned} dx_1(t) &= \mu_1(t, x_1, x_2)dt + \sigma_1(t, x_1, x_2)dW_{x_1}^{\mathbb{Q}}(t) \\ dx_2(t) &= \mu_2(t, x_1, x_2)dt + \sigma_2(t, x_1, x_2)dW_{x_2}^{\mathbb{Q}}(t) \\ dW_{x_1}^{\mathbb{Q}}(t)dW_{x_2}^{\mathbb{Q}}(t) &= \rho(t, x_1, x_2)dt, \end{aligned}$$

and the pricing problem (on the set $\{\tau^ > t\}$)*

$$\tilde{V}(t, T) = \mathbb{E}^{\mathbb{Q}} \left[\tilde{\eta}(t, T)\tilde{g}(x_1(T), x_2(T)) + \int_t^T \tilde{\eta}(t, u)\tilde{h}(u, x_1(u), x_2(u))du \middle| \mathcal{F}(t) \right],$$

⁷⁷notice that the (square) intensity's lowest value on the grid has to be > 0 .

where

$$\begin{aligned}\tilde{g}(x_1(T), x_2(T)) &= g(x_1(T), e^{x_2(T)}) \\ \tilde{\eta}(t, T) &= \exp\left(-\int_t^T (x_1(u) + e^{x_2(u)}) du\right) \\ \tilde{h}(u, x_1, x_2) &= e^{x_2(u)} Z(u, x_1, e^{x_2(u)}) + c(u, x_1, e^{x_2(u)}).\end{aligned}$$

Then $\tilde{V}(t, T)$ solves

$$\begin{aligned}\frac{\partial \tilde{V}(t, x_1, x_2)}{\partial t} + \mathcal{A}\tilde{V}(t, x_1, x_2) + e^{x_2} Z(t, x_1, e^{x_2}) + c(t, x_1, e^{x_2}) \\ - (x_1 + e^{x_2})\tilde{V}(t, x_1, x_2) = 0,\end{aligned}\tag{III.2.6}$$

where

$$\begin{aligned}\mathcal{A} &= \mu_1(t, x_1, x_2) \frac{\partial}{\partial x_1} + \mu_2(t, x_1, x_2) \frac{\partial}{\partial x_2} + \frac{1}{2} \sigma_1^2(t, x_1, x_2) \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} \sigma_2^2(t, x_1, x_2) \frac{\partial^2}{\partial x_2^2} \\ &+ \sigma_1(t, x_1, x_2) \sigma_2(t, x_1, x_2) \rho(t, x_1, x_2) \frac{\partial^2}{\partial x_1 \partial x_2}.\end{aligned}$$

III.2.1 Applications of the fundamental PDE

The main result of the previous section was a general pricing PDE (III.2.2) for securities subject to default risk governed by a certain stochastic intensity process. We also provided useful variants of this PDE customized for particular problems. It remains to demonstrate the power and the great flexibility of the PDE on selected problems.

In what follows we adopt the following notation which already appeared in the previous section, yet without further comments. $V = V(t, r, \gamma)$ is the value of the financial contract in question *in terms of the non-transformed, original variables* r, γ . For exact application-ready form of the PDE it is, however, in almost all cases better to work with the transformed variables and so often we use e.g. $x_2 = \sqrt{\gamma}, x_2 = \ln \gamma$ instead of just $x_2 = \gamma$. The same holds for any transformation of the risk-free rate r .⁷⁸ We denote $\tilde{V}(t, x_1, x_2)$ the value of the financial contract *in terms of the transformed variables* $x_1 = f_1(r), x_2 = f_2(\gamma)$, where $f_i, i = 1, 2$ are some transformation functions. Notice that $V(t, r, \gamma) = \tilde{V}(t, x_1, x_2)$ so both variables V and \tilde{V} refer to the *same value* of the financial instrument and they differ only in that they are related to the (non)transformed processes. Finally, if $x_1 = r, x_2 = \gamma$ then there is no transformation of variables and we often write $V(t, x_1, x_2)$ instead of $V(t, r, \gamma)$. If x_1, x_2 refer to transformed or non-transformed variables will always be clear from the context.

III.2.1.1 Pricing of a corporate bond and its embedded option

The project is to price a corporate bond with an embedded option and extract the value of the optionality in the setup where the interest rates and the default intensity process of the

⁷⁸we have not provided any PDE where a transformation of r would be included but the form of the PDE would be more or less an analogy of the outlined 'intensity-transformed' PDEs.

issuer are both stochastic and possibly correlated processes. The central idea is to price a defaultable bond with (i) no optionality and (ii) with the optionality and obtain the value of the embedded option as their difference. Thus, the problem of pricing the embedded option breaks down into pricing of two defaultable bonds and the value of the embedded option is obtained as a by-product. The flexibility of the pricing PDE allows for rich features to be built in. For example, the bond can be a floater paying some floating rate + margin. Also the recovery can be subject to various recovery policies. In the illustration below, we assume the case of fixed coupon bond which is for the pricing of embedded options more attractive than the floater because it is more sensitive to fluctuations of the risk-free rate. We also show that the embedded put option is more valuable in a corporate (defaultable) bond than the same embedded option in otherwise similar but default-free bond. This makes sense because a rational investor would put the bond if its credit is deteriorating, this deterioration however can not happen in the case of risk-free bond whose value is entirely determined by the interest rates.

Assume a defaultable puttable bond $V(\cdot, T)$ paying continuously coupons at rate c . The rate can be set up so that effectively it is equivalent to some discretely paid coupon rate (see e.g. (Radová et al., 2009)). We assume that upon default the bond recovers δ of its notional. The bond is puttable at a pre-specified price H at time $T^* < T$. If the short-rate r follows the Vasicek dynamics (Vasicek, 1977) and the hazard rate γ follows CIR, then according to the Theorem III.2.2 the bond $V = \tilde{V}$ prior to default satisfies

$$\begin{aligned} \frac{\partial \tilde{V}(t, x_1, x_2)}{\partial t} + \mathcal{A}\tilde{V}(t, x_1, x_2) + \left((x_2)^2 \delta + c \right) - \left(x_1 + (x_2)^2 \right) \tilde{V}(t, x_1, x_2) &= 0 \\ \tilde{V}(T, x_1, x_2) = \tilde{g}(x_1, x_2) = g(x_1, x_2^2) &= 1, \end{aligned} \tag{III.2.7}$$

where the infinitesimal generator \mathcal{A} takes the particular form

$$\begin{aligned} \mathcal{A} &= \kappa_1(\theta_1 - x_1) \frac{\partial}{\partial x_1} + \frac{1}{2} \left(\frac{1}{x_2} \left(\kappa_2 \theta_2 - \frac{1}{4} \sigma_2^2 \right) - \kappa_2 x_2 \right) \frac{\partial}{\partial x_2} \\ &+ \frac{1}{2} \sigma_1^2 \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} \frac{1}{4} \sigma_2^2 \frac{\partial^2}{\partial x_2^2} + \frac{1}{2} \sigma_1 \sigma_2 \rho \frac{\partial^2}{\partial x_1 \partial x_2}. \end{aligned}$$

The processes x_1, x_2 are specified as $x_1 = r, x_2 = \sqrt{\gamma}$, where r is driven by Vasicek and γ follows CIR (thus x_2 follows the tempered CIR (III.2.3) (see Lemma III.2.1)).

Let us consider the following setup. For the risk-free rate we assume $r = 0.05, \kappa_1 = 0.3, \theta_1 = 0.06, \sigma_1 = 0.015$. The credit risk of the issuer is subject to a CIR process with $\gamma = 0.2, \kappa_2 = 0.4, \theta_2 = 0.25, \sigma_2 = 0.03$ and recovery $\delta = 0.4$ paid at the time of default. The bond has $T = 5$ years to maturity, coupon rate $c = 0.04$ and is puttable at $T^* = 2.5$ at the price $H = 1$. In order to analyze the effect of the risk-free rate and the credit risk, we assume 4 hypothetical bonds to be traded:

- A : a bond with no default risk ($\gamma \equiv 0$) and no embedded option,
- B : a bond with no default risk ($\gamma \equiv 0$) and with an embedded option (puttable at $H = 1$ at time $t = T^*$),
- C : a bond subject to default risk and no embedded put option,

D : a bond subject to default risk and with an embedded put option (puttable at $H = 1$ at time $t = T^*$).

Solving the PDE (III.2.7) for all the four claims A–D as specified above gives the results shown in the Table 2. The bond A has no credit risk and no optionality. Therefore its

bond	defaultable	puttable	price	credit risk impact	embedded option
A	no	no	0.9372	0.0000	0.0000
B	no	yes	0.9711	0.0000	0.0339
C	yes	no	0.5966	-0.3406	0.0000
D	yes	yes	0.7394	-0.2317	0.1428

Table 2: Summary table for bonds A, B, C and D and their decomposition.

price represents purely the time-value of the bond. Adding put-optionality to A gives the bond B and we see the price has grown slightly to 0.9711 which is an increase of only 0.0339 in the value of the bond. In the case of the bond C (defaultable and no optionality) the value of only 0.5966 is observed. This is because the bond carries significant credit risk and the holder has no right to put it. Compared to its default-free counterpart (A), there is a -0.3406 drop in the value. Finally, the bond D represents the case of C with embedded put option. By taking difference of the price of D and C we obtain the value of the embedded put, which is 0.1428. Comparing the value 0.0339 of the option embedded in the default-free bond and the value 0.1428 of the option embedded in a defaultable bond we conclude that the embedded option is much more valuable if being part of a risky bond. Hence the credit risk of the issuer has a significant impact on the value of the embedded (put) option.

Although the pricing surface for the embedded option can not be captured directly, by differencing the pricing surfaces of the bond C and D we obtain the pricing surface of the embedded put. This is shown in Figure 8. The picture also explains how the value of the embedded option behaves with respect to r and γ (or its square root). We can observe the following effects. If r is low and γ is low, then the option is quite valuable because the mean reversion level θ_1 is above r and so the interest rate tends to rise in the future, and it is relatively unlikely that the bond defaults until T^* and hence the option could be struck. If r is low but γ is high then the option loses its value because it is unlikely that the company survives until T^* . The price of the option is even lower when r is high and γ is high. In that scenario it is unlikely that the company survives until T^* and moreover the discounting rate is high (lower PV). We can see the sweet spot is somewhere at the coordinate $r = 0.1$ and when also the intensity γ is low. The idea is that γ is likely to increase in the future ($\theta_2 > \gamma(0)$) which might reduce the bond price in the future and the risk-free rate is still quite high. This is the ideal scenario for the option to be put.

Remarks. If the underlying bond were a floater paying $r(t) + s$, where s is some fixed spread, it would only suffice to replace c by $(x_1 + s)$ in the PDE (III.2.7). The scheme is also trivially extensible to multiple exercise dates (and various exercise strikes), not necessarily exactly only one date T^* is the limit. The value of every additional embedded

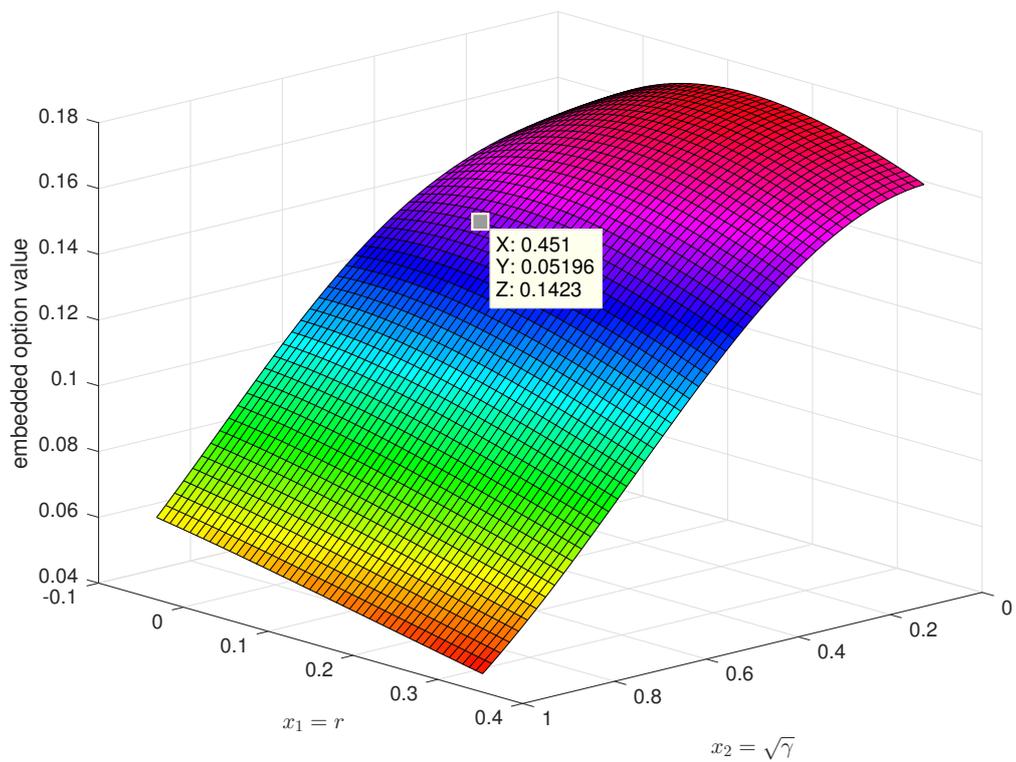


Figure 8: Pricing surface of the put option embedded in the corporate bond $V(t, T)$ at $t = 0$. The spatial variables are the risk-free rate r and the square root of the intensity $\sqrt{\gamma}$.

option would as a result decrease the value of the embedded put option we calculated. This is because the option considered in this example could become suboptimal to be exercised if other 'more attractive' options can be struck. Clearly, purely analogously we could also analyze the impact of an embedded call which in turn would reduce the value of the bond.

III.2.1.2 Pricing the counterparty risk in an IRS contract

Despite the current market standard that IRS deals are collateralized, frequently the collateralization might not be perfect or there might even not be any collateralization at all. The objective of this pricing case is to investigate the counterparty credit risk aspect of such a deal, in view of the main pricing PDE in the Theorem III.2.1. Here, we are fully aware that the outcome is approximative, yet the bias is rather mild. In spite of existence of a semi-closed formula for IRS CVA (see e.g. the recent article (Cerny and Witzany, 2014)) our objective is to connect the pricing and CVA with collateralization whose marginal cases are the perfect collateralization and no collateralization. The collateralization is represented by a cash collateral account U .

This setup, in turn, has a special effect on the recovery function Z . If a recovery *rate* is pre-specified, then upon default it is typically assumed that a given portion of the deal's MtM is recovered, regardless of the MtM itself (given it is positive). Under the *collateral* setup, the recovery rate is an implicit function of both the collateral balance U and the MtM of the IRS. For low (but positive) MtM, a fixed amount of collateral U shall fully cover the exposure but for high exposures the same amount of the collateral is insufficient and the incurred loss increases, if default event of the counterparty occurs. In this stylized example, to make the illustration more informative, we place ourselves into the position of receiver IRS whose position in the IRS has a very high market value in this stylized setup.

Assume a receiver IRS deal where the coupon is received at rate K and paid at rate $r(t)$. Both rates are assumed to be under the continuous compounding, coupons are assumed to be paid continuously and also the rate r resets continuously in time.⁷⁹ Assume such IRS being the contingent claim in question. Then this contingent claim bears a coupon ($K - r(t)$) (as seen from the fix receiver) until the time $\min[\tau^*, T]$, where τ^* is the counterparty's default time as usual. Assume the rate r follows the Hull-White's dynamics

$$dr(t) = \kappa(\theta(t) - r(t))dt + \sigma dW^{\mathbb{Q}}(t).$$

By re-transforming this dynamics into the dynamics of $x(t) = r(t) - f(0, t)$ (see the Appendix A.4) the dynamics of x is now

$$dx(t) = (y(t) - \kappa x(t))dt + \sigma dW^{\mathbb{Q}}(t), x(0) = 0,$$

where $y(t) = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t})$. The bond solution in terms of x is then

$$\begin{aligned} B(t, T) &= \frac{B(0, T)}{B(0, t)} \exp\left(-x(t)G(t, T) - \frac{1}{2}y(t)G(t, T)^2\right) \\ G(t, T) &= \frac{1}{\kappa}(1 - e^{-\kappa(T-t)}). \end{aligned}$$

⁷⁹this is the approximation that we referred to earlier.

The value of the *default-free* receiver IRS in terms of x such that $B(t, T) \equiv B(t, T, x)$ is then⁸⁰

$$V_{\text{rf}}(t, T) \approx K\Delta \sum_{i=1}^n B(t, t + i\Delta) - 1 + B(t, T), \Delta = \frac{T-t}{n}, n \rightarrow \infty.$$

To incorporate the recovery policy, let the collateral cash account $U = U(t, x_1, x_2)$ be some nonrandom function of time t and the state variables x_1, x_2 (these can be the transformed variables). U models the amount of collateral posted by the counterparty to (possibly only partially) cover the exposure. Let us now define $x_1(t) = x(t) := r(t) - f(0, t)$ and $x_2 = \sqrt{\gamma}$. Then the value of the *vulnerable* IRS $V(t, r, \gamma) = \tilde{V}(t, x_1, x_2)$ which is subject to counterparty default risk solves

$$\begin{aligned} \frac{\partial \tilde{V}(t, x_1, x_2)}{\partial t} + \mathcal{A}\tilde{V}(t, x_1, x_2) + \left((x_2)^2 \tilde{Z}(t, x_1, x_2) + \tilde{c}(t, x_1, x_2) \right) \\ - \left(x_1 + f(0, t) + (x_2)^2 \right) \tilde{V}(t, x_1, x_2) = 0, \text{ (III.2.8)} \end{aligned}$$

with the terminal condition, coupon function and recovery function

$$\begin{aligned} \tilde{V}(T, x_1, x_2) &= 0 \\ \tilde{c}(t, x_1, x_2) &= c(t, x_1 + f(0, t), (x_2)^2) = K - (x_1 + f(0, t)) \\ \tilde{Z}(t, x_1, x_2) &= \min[V_{\text{rf}}(t, T), U(t, x_1, x_2)] \mathbf{1}_{\{V_{\text{rf}}(t, T) > 0\}} + V_{\text{rf}}(t, T) \mathbf{1}_{\{V_{\text{rf}}(t, T) < 0\}}, \end{aligned}$$

respectively. To this end, the infinitesimal generator \mathcal{A} reads

$$\begin{aligned} \mathcal{A} &= (y(t) - \kappa_1 x_1) \frac{\partial}{\partial x_1} + \frac{1}{2} \left(\frac{1}{x_2} \left(\kappa_2 \theta_2 - \frac{1}{4} \sigma_2^2 \right) - \kappa_2 x_2 \right) \frac{\partial}{\partial x_2} \\ &+ \frac{1}{2} \sigma_1^2 \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} \frac{1}{4} \sigma_2^2 \frac{\partial^2}{\partial x_2^2} + \frac{1}{2} \sigma_1 \sigma_2 \rho \frac{\partial^2}{\partial x_1 \partial x_2}, \end{aligned}$$

where y is a deterministic function of time depending only on the fixed parameters σ_1, κ_1 related to the Hull-White process.

We assume being placed in a USD fixed-income market as of Feb 15, 2016. The project is to extract the counterparty's credit risk from an IRS deal which has already been entered into some time ago as fixed receiver at the coupon rate $K = 0.05$ and the IRS is maturing in $T = 5$ years. In order to obtain a good market match, we have calibrated the HW model using the approach described in the Appendix A.4 to a set of ATM swaptions whose underlying IRS mature in $T = 5$ years (26 swaption quotes in total). This gave us the HW parameters $\kappa_1 = -0.0998, \sigma_1 = 0.0068$. The quality of the fit is obvious from the Figure 9 and it is easy to see that the fit is almost perfect, which proves HW being a good choice for the dynamics of the interest rate r . The counterparty's default intensity is assumed to be driven by CIR, specified by $\gamma = 0.17, \kappa_2 = 0.1, \theta_2 = 0.23, \sigma_2 = 0.04$. We assume that according to the credit support annex (CSA), the counterparty's exposure

⁸⁰notice $\lim_{n \rightarrow \infty} \left(\Delta \sum_{i=1}^n B(t, t + i\Delta) \right) = \int_t^T B(t, u) du$ but for the practical implementation it is better to keep the 'discrete-form' expression and select some judicious value of n . Usually $n = 50$ is sufficient.

threshold χ is 0.1.⁸¹ This means that if the value of the default-free value of the IRS is, say 0.15, the counterparty needs to maintain its collateral account at 0.05 (and the exposure 0.1 remains uncovered). In this view, $U(t, x_1, x_2) = \max[V_{\text{rf}}(t, T) - \chi, 0]$. Hence, the threshold represents the maximum uncovered exposure or the hypothetical maximum loss due to the default of the counterparty.

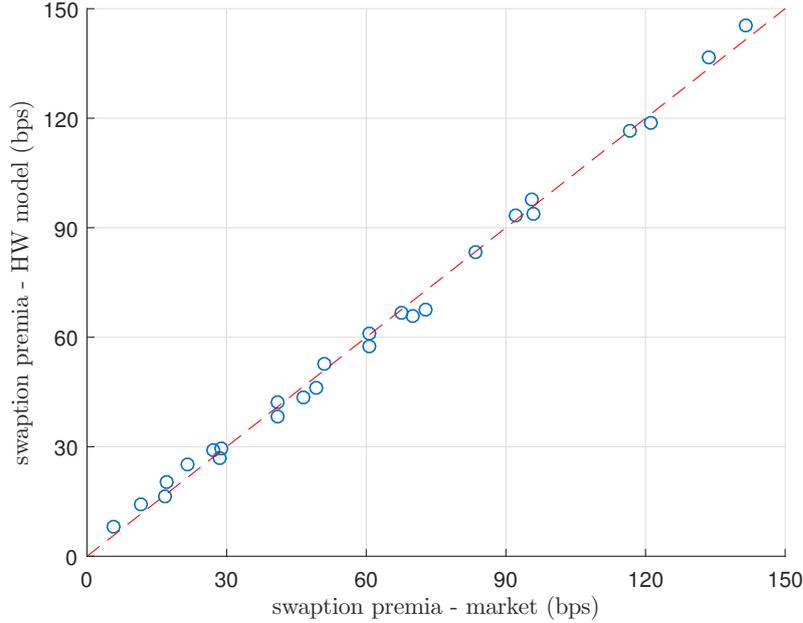


Figure 9: Scatterplot of calibrated HW model swaption premiums vs market swaption premiums in the USD fixed income market as of Feb 15, 2016. All swaption quotes were ATM, with the underlying IRS maturing within 5 years. The dashed red line is a perfect fit line.

After solving the PDE⁸² (III.2.8), we have found that the value of such a partially collateralized receiver IRS position is 0.1450. Conducting analogous computation for the same but default-free deal gives the value of the receiver IRS 0.1889. Therefore we conclude that the amount of the counterparty credit risk in the deal is relatively high, about 439 bps. This is also fundamentally caused by the fact that the receiver IRS position is quite valuable at the time of valuation (notice the coupon received is $K = 0.05$ and the paid floating rate is close to 0). Also, the collateralization is far from being optimal, the threshold $\chi = 0.1$ reflects a relatively high tolerance to the counterparty's credit risk. To provide a more detailed look at the credit risk, we portray the credit risk profile in the Figure 10. The picture has been obtained by subtracting the pricing surface of the receiver IRS (with credit risk given by the threshold $\chi = 0.1$) from the receiver IRS surface with no credit

⁸¹for simplicity we assume the CSA is assumed to be on a transaction level. In more general cases the CSA is related to a portfolio of transactions.

⁸²with custom designed, highly optimized functions we solve the PDE in < 0.15 s given 80 spatial points into both x_1, x_2 directions. The optimal numerical approaches are described in the Appendix C.

risk (this has been achieved by setting $\chi = 0$ which effectively means that the IRS is perfectly collateralized at any time). Any point on this surface shows by how much we shall decrease the risk-free deal to obtain the true value of the IRS in which the counterparty is allowed to default. Hence, higher values point to higher counterparty credit risk in the deal which results into lower values of the credit risk adjusted IRS. It is easy to see that the fundamental PDE was perfectly able to incorporate the interplay between the credit risk and market value of the IRS. Apparently, any area on this profile for $x_1 > 0.05$ has value 0. This is because of the fact that for high values of the risk-free rate the value of the receiver IRS position is low or negative and therefore in such scenarios there is no counterparty credit risk.⁸³ The amount of credit risk, however, increases if the interest rate r (or x_1 , equivalently) is low. For low r the receiver IRS has a high exposure to the counterparty and bears credit risk according to the threshold χ and the default intensity γ . As we have $\chi = 0.1$ fixed, the picture provides the amount of credit risk for various values of γ (or its square root, respectively). Thus for the scenarios of very low r (high value of the receiver IRS) the credit risk increases with increasing default intensity γ . Hence, the Figure 10 exhaustively explains all the factors that determine the counterparty credit risk of the given deal. We finalize the analysis with one last supplementary picture. Figure

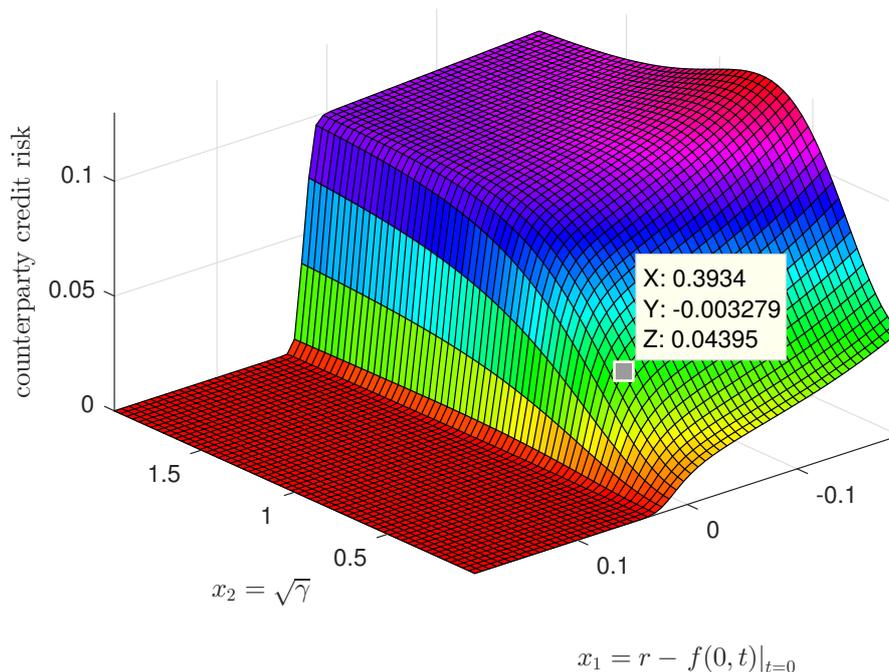


Figure 10: Surface of counterparty credit risk in the analyzed IRS deal, as seen from the receiver IRS perspective as of Feb 15, 2016.

11 shows the value of the receiver IRS position in question as a function of the threshold χ specified in the CSA. This can be particularly useful for optimization of CSA contracts

⁸³should the counterparty default in such a scenario, we suffer no *credit* loss because the value of the contract was negative anyway.

where the officer who is in charge of the CSA management should cautiously assess all aspects of the deals that are already running or about to be opened in the future. If χ is low, then the counterparty must fully collateralize its position of negative value (from the counterparty's perspective), this means the deal is made risk-free regardless of the riskiness of the counterparty and so the value of the deal for $\chi = 0$ equals the default-free value of the receiver IRS (0.1889). The curve depicting the value of the IRS position, however, decreases sharply with increasing threshold χ . The justification for this effect is simple. As the receiver IRS position has a high MtM, counterparty's default would result into losses. In turn, any supplied collateral would be almost certainly used to eliminate the loss resulting the the default. Hence, increasing χ (=less collateral supplied) almost surely results into some default-loss and so the deal's counterparty credit risk-adjusted value is very sensitive to the threshold χ at the short end. The curve stops to decrease around $\chi = 0.25$. The argument is that it is unlikely in the future that the exposure in the IRS contract goes beyond 0.25 and so the maximum hypothetical loss due to default is limited and increasing the threshold will not increase the possible loss.

In the Section III.2.2.1 we propose a PDE for a direct calculation of the CVA such that the CVA is the exclusive outcome of the PDE and the CVA surface does not need to be calculated using the differencing principle which has been applied here.

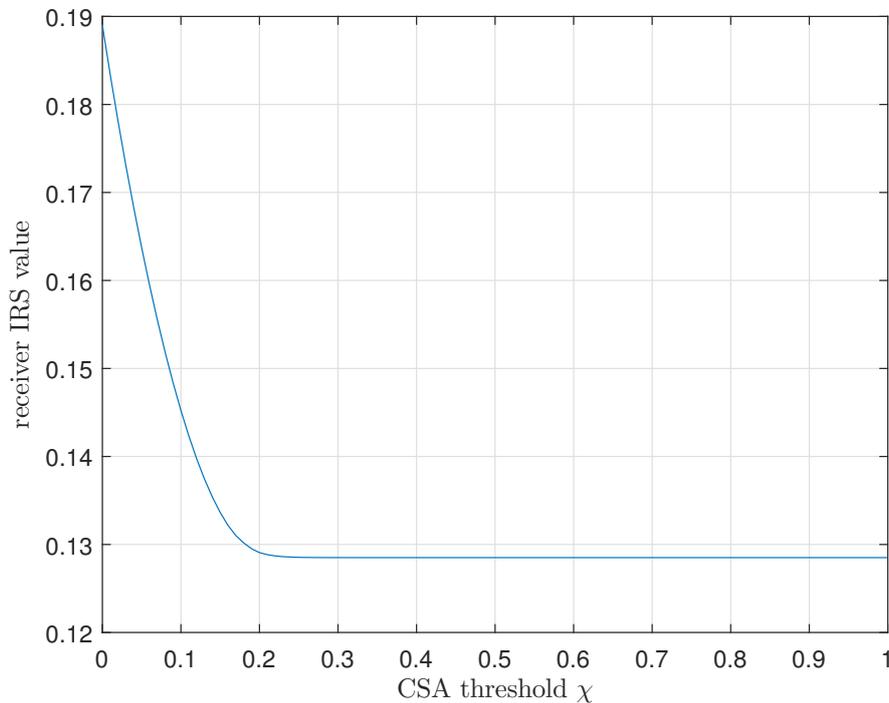


Figure 11: Value of the receiver IRS position as a function of the CSA threshold χ .

Remarks. In many applications we might want to price an *amortizing* IRS whose counterparty is subject to default risk. This common feature is not difficult to be implemented in the model. Basically, it would only require to replace in (III.2.8) the 'coupon' function

$\tilde{c}(t, x_1, x_2) = (K - (x_1 + f(0, t)))$ with $\tilde{c}(t, x_1, x_2) = L(t, x_1, x_2) (K - (x_1 + f(0, t)))$ where L is the outstanding notional of the IRS at time t (it can furthermore be dependent of x_1, x_2). Moreover L does not need to be a continuous function, it might be e.g. specified as a (not necessarily decreasing) step function

$$L(t, x_1, x_2) = L(t) = \begin{cases} 1.0 & \text{if } t \leq t_1 \\ 0.7 & \text{if } t_1 < t \leq t_2, \\ 0.9 & \text{if } t_2 < t \leq T \end{cases}$$

where t_1, t_2 are some fixed, pre-specified dates of the IRS amortization.

III.2.1.3 Calibration, pricing and valuation of CDS

In the previous sections, it was assumed that the parameters of the dynamics of the intensity are given. In practice, however, these parameters have to be estimated. In this section we address the calibration of the stochastic intensity parameter set $\Theta_2 = (\gamma, \kappa_2, \theta_2, \sigma_2)$ if CIR intensity process is assumed. In this part we besides CIR also consider Black-Karasinski with fixed parameters, i.e.

$$dx_2(t) = \kappa_2(\theta_2 - x_2(t))dt + \sigma_2 dW(t), x_2(t) = \ln \gamma(t), \quad (\text{III.2.9})$$

because it is a log-normal model and so the intensity can never go below 0.⁸⁴ The parameter set is the same as in the CIR case. Although Black-Karasinski does not allow for closed-form solutions, it is not a reason for being concerned as the only form we are interested in is its differential (III.2.9) which has a very convenient form. We further combine the question of calibration with the CDS pricing.

CDS is probably the best direct measure of the credit risk of a selected counterparty which means it serves as a reliable and dedicated source of credit information without any noise which could be present in e.g. bonds, where one has to apply some estimate of recovery to price the bond correctly (CDS contracts are traded under explicitly specified recovery δ). Since a CDS is a contract subject to credit risk, it will also satisfy the fundamental PDE (III.2.2). The natural approach to calibration is to collect a term-structure of CDS quotes, value those contracts in the PDE (III.2.2) such that the value of every such a contract is close to 0 by iterating the parameter set Θ_2 . As the primary focus is on the credit in the CDS pricing, we will assume that r is a fixed constant. This simplification is not necessary and easily can be relaxed. However, as the contribution of the randomness in r to the CDS price/value is negligible.

Let us assume at time t a set of $i = 1, \dots, n$ CDS contracts V_i with tenors τ_i written on the same name, each maturing at T_i .⁸⁵ The CDS spread of i th contact which makes $V_i = 0$ is labelled as K_i . We also assume for every CDS contract some constant risk-free rate r_i .⁸⁶ If we place ourselves into the position of the *protection seller*, we receive in the

⁸⁴notice that CIR can go below 0 if the Feller condition is not satisfied.

⁸⁵ T_i is not necessarily $t + \tau_i$. For example when a 5Y CDS contract has been entered into some time ago and has 3 years of life remaining at the current time $t = 0$, then $T_i = 3$ but $\tau_i = 5$ is still a constant 'parameter'.

⁸⁶this quantity is related to the maturity T_i .

continuous world in the notation of the PDE (III.2.2) a coupon $c_i = K_i$ until $\min[\tau^* - , T_i]$. Should the underlying entity default, we are obliged to *pay* $(1 - \delta_i)$, where δ_i is the quoted recovery in i th contract. There is also no terminal payment at time T_i (except if there were a default at exactly T_i).

This means that the value (not the price) of the contract $\tilde{V}_i(t, \ln \gamma) = V_i(t, \gamma)$ on the set $\{\tau^* > t\}$ in terms of the log intensity $x_2 = \ln \gamma$ satisfies the PDE

$$\begin{aligned} \frac{\partial \tilde{V}_i(t, x_2)}{\partial t} + \mathcal{A} \tilde{V}_i(t, x_2) + (-e^{x_2}(1 - \delta_i) + K_i) - (r_i + e^{x_2}) \tilde{V}_i(t, x_2) &= 0 \\ \tilde{V}_i(T_i, x_2) &= 0, \end{aligned} \tag{III.2.10}$$

given the operator

$$\mathcal{A} = \begin{cases} e^{-x_2} \left(\kappa_2(\theta_2 - e^{x_2}) - \frac{1}{2} \sigma_2^2 \right) \frac{\partial}{\partial x_2} + \frac{1}{2} \sigma_2^2 e^{-x_2} \frac{\partial^2}{\partial x_2^2} & \text{for (log) CIR} \\ \kappa_2(\theta_2 - x_2) \frac{\partial}{\partial x_2} + \frac{1}{2} \sigma_2^2 \frac{\partial^2}{\partial x_2^2} & \text{for Black-Karasinski} \end{cases}.$$

Knowing how to compute the value of the CDS, we can calibrate the model by finding optimal parameters Θ_2 by requiring that the optimal parameter set makes the values of all the quoted contracts as close to zero as possible. The estimator $\hat{\Theta}_2$ of Θ_2 is then defined as

$$\hat{\Theta}_2 = \arg \min_{\Theta_2} [\varepsilon(\Theta_2)], \varepsilon(\Theta_2) = \sum_{i=1}^n \left| \tilde{V}_i(t | \Theta_2) \right|, \tag{III.2.11}$$

which captures the idea that the value of a CDS contract must be zero at the inception of the CDS deal if contracted under a fair spread. To test the quality of the setup, we have collected quotes of CDS written on 10 large banks. These quotes are shown in the Table 3. If the intensity models (CIR, Black-Karasinski) can match these prices correctly by making every contract running under the quoted spread zero-valued, then the fit is perfect. After we ran the calibration described above, we have obtained parameters summarized in the Table 4 for CIR and in the Table 5 for Black-Karasinski. The average valuation error $\bar{\varepsilon}$ indicates the fits are nearly perfect for all the names and hence both CIR and Black-Karasinski seem to be well-suited models for capturing the default intensity γ . The Black-Karasinski is even so good that its errors are approximately only one half of the errors produced by the CIR.

Having calibrated the models, CDS pricing and valuation is simple. Both can be done by means of (III.2.10). Let us assume being a protection seller in a CDS contract written on the counterparty Morgan Stanley (MS) which was entered into at $K = 0.02$ some time ago and that the contract has 13 years remaining. Then according to the (III.2.10) the value of this position is equal to 0.0380 or 380 bps in CIR and 0.0365 or 365 bps in Black-Karasinski, with respect to the notional of the deal. The resulting valuation surfaces are shown in the Figure 12 (log CIR) and Figure 13 (Black-Karasinski). As the 13Y MS CDS is according to the Table 3 somewhere between 153 bps (10Y deal) and 173 bps (20Y deal) bps but in the deal contracted earlier we receive 200 bps, the mark to market value must be positive for the protection seller. We can also use the

CDS tenor (τ_i)	MS	GS	DB	CITI	WF	UBS	CS	BoA	JPM	HSBC
6M	28	37	157	34	16	31	104	35	30	48
1Y	48	46	165	49	20	39	111	50	37	56
2Y	68	59	175	62	30	50	120	66	48	70
3Y	81	77	184	74	42	58	129	77	53	85
4Y	94	97	191	86	52	70	141	89	63	101
5Y	113	115	196	103	64	78	151	105	78	117
7Y	138	137	207	121	78	95	168	126	97	141
10Y	153	154	214	139	94	108	180	142	117	164
20Y	173	175	215	168	113	115	201	171	143	190
30Y	179	181	215	176	119	117	207	180	151	197

Table 3: Set of CDS quotes (in bps) for various names and tenors as of March 25, 2016.

$\hat{\Theta}_2$	MS	GS	DB	CITI	WF	UBS	CS	BoA	JPM	HSBC
γ	0.0026	0.0029	0.0235	0.0040	0.0005	0.0027	0.0146	0.0050	0.0025	0.0008
κ_2	0.3133	0.2951	0.5324	0.2027	0.1961	0.4393	0.2203	0.1837	0.1743	0.2750
θ_2	0.0372	0.0379	0.0383	0.0397	0.0283	0.0224	0.0423	0.0419	0.0363	0.0436
σ_2	0.0796	0.0728	0.0773	0.0473	0.1036	0.0804	0.0427	0.0623	0.0673	0.0949
$\bar{\varepsilon}$	10.9	11.0	9.0	4.0	3.7	14.5	4.0	5.7	9.9	7.8

Table 4: Calibrated parameters of the CIR intensity process from CDS quotes as of March 25, 2016. The average error $\bar{\varepsilon}$ (in bps) represents ε in (III.2.11) divided by the number of contracts in the dataset ($n = 10$). It is the average error in *value* of a CDS on a given name when valued via the PDE (III.2.10) compared to the hypothetical zero fair value.

$\hat{\Theta}_2$	MS	GS	DB	CITI	WF	UBS	CS	BoA	JPM	HSBC
γ	0.0058	0.0034	0.0250	0.0061	0.0019	0.0041	0.0159	0.0063	0.0043	0.0066
κ_2	0.1278	0.5079	0.1127	0.2581	0.4035	0.0960	0.1566	0.2264	0.1194	0.1183
θ_2	-4.0707	-3.6050	-3.6254	-3.5256	-4.0432	-5.3449	-3.3792	-3.6037	-4.0337	-3.6459
σ_2	1.0469	0.8313	0.5913	0.6370	0.7986	1.1437	0.5436	0.7369	0.9196	0.9344
$\bar{\varepsilon}$	6.9	5.6	3.2	3.9	2.3	5.7	3.8	5.8	4.0	2.9

Table 5: Calibrated parameters of the Black-Karasinski intensity process from CDS quotes as of March 25, 2016. The average error $\bar{\varepsilon}$ (in bps) represents ε in (III.2.11) divided by the number of contracts in the dataset ($n = 10$). It is the average error in *value* of a CDS on a given name when valued via the PDE (III.2.10) compared to the hypothetical zero fair value.

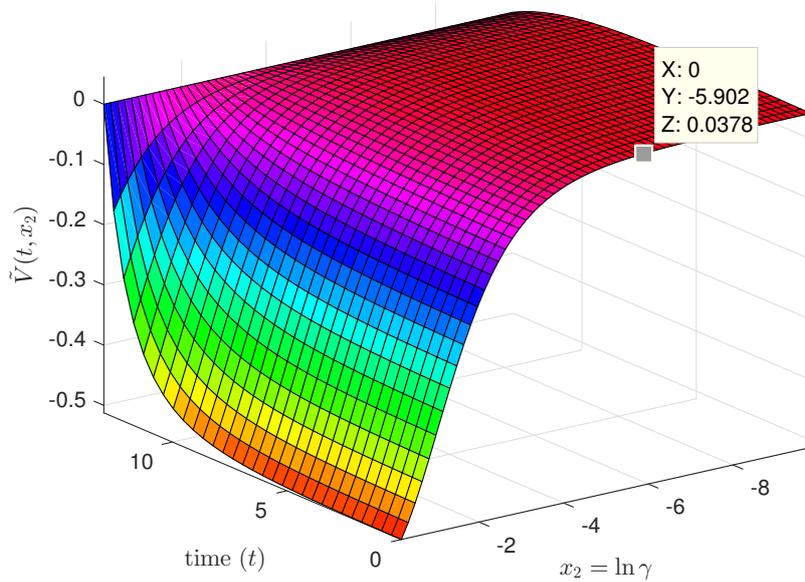


Figure 12: Log CIR valuation surface of 13Y CDS written on Morgan Stanley as seen from the protection seller, entered into at $K = 0.02$. The valuation is as of March 25, 2016. The solution lies at $x_2 = \ln(\gamma) = \ln(0.0026) = -5.9522$ because Morgan Stanley's intensity was estimated to be $\gamma = 0.0026$ as shown in the Table 4.

PDE (III.2.10) to calculate the fair spread which would make 13Y deal to have zero value as $K^* = 0.0163$ or 163 bps in CIR and $K^* = 0.0165$ or 165 bps in Black-Karasinski which both are almost perfectly compatible theoretical 13Y CDS quotes with the Table 3. Notice that the PDE (III.2.10) does not allow to directly express the fair spread K_i but rather *connects* the spread K_i with the value of the deal \tilde{V} . Therefore K_i in terms of (III.2.10) needs to be found by a numerical search over various values of K_i and the value of $K_i = K_i^*$ which makes the value of the deal 0, is the correct fair spread. This critical idea will eventually play a central role in CDS *option* pricing in the Section III.2.1.4.

Remarks. Surprising results in the quality of fit to the CDS quotes were obtained when we modified the Black-Karasinski's diffusion part to CEV-like such that log-intensity SDE was specified as

$$d \ln \gamma(t) = \kappa_2(\theta_2 - \ln \gamma(t))dt + \gamma(t)^\alpha \sigma_2 dW(t), 0 \leq \alpha \leq 1.$$

For $\alpha = 0$ the model coincides with the Black-Karasinski. We observed a particularly good performance with $\alpha = 1/3$. Under this specification the model managed to outperform even the Black-Karasinski's result in the Table 5. This would indicate that the behaviour of the default intensity γ follows some rare process such that the distribution of γ is non-standard.

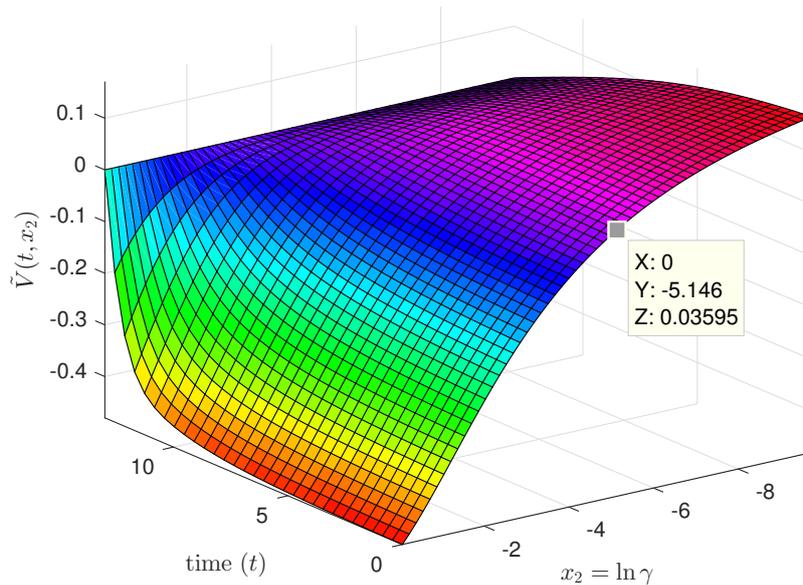


Figure 13: Black-Karasinski valuation surface of 13Y CDS written on Morgan Stanley as seen from the protection seller, entered into at $K = 0.02$. The valuation is as of March 25, 2016. The solution lies at $x_2 = \ln(\gamma) = \ln(0.00578) = -5.1534$ because Morgan Stanley’s intensity was estimated to be $\gamma = 0.00578$ as shown in the Table 4.

III.2.1.4 Pricing of CDS options

In the previous Section III.2.1.3 we have demonstrated how to price the CDS contracts using the PDE methods in the stochastic intensity framework. In this section we push the ideas even further and show how the methodology can be extended for pricing of CDS options. A *CDS option* is a contract allowing to enter a CDS contract in a given future date at a specified premium (coupon) rate. Hence, it is no different from e.g. stock option only with the difference that the underlying is a CDS contract which is a substantially more complicated asset than a stock.

In turn, the pricing of options on CDS is certainly a challenging task. A classical methodology such as (Hull and White, 2003) are assuming some given deterministic inputs which we consider as a limitation as the markets are not behaving deterministically. The main complication in the CDS option pricing is that the dynamics of the underlying CDS quotes are difficult to be described model-wise and apart from that the CDS curve is tradeable so the modeled dynamics of the CDS quote would have to ensure that there is no arbitrage.

In the case of a stock its motion is approximately ‘geometric’ and so there is a hint of how to model its dynamics. This is a much more challenging issue for the CDS, yet for the CDS options the dynamics of CDS is a critical component. We show that the CDS option prices can be efficiently computed within the PDE framework in the way that completely eliminates any model-dependence on the problematic dynamics of the CDS quotes. It

is also easy to see that the pricing problem of CDS options in the stochastic intensity framework is technically prohibitive even when using MC methods so the PDE approach is a reasonable technique for this problem. To illustrate the complexity of the problem in the MC framework, if we wanted to price a CDS option, we would have to simulate γ until the CDS option's maturity T . There the MC would have to be run again to price the CDS given $\gamma = \gamma(T)$ and when having the price of the CDS we would compare this value with CDS option strike K and discount back to 0. This makes the problem solvable in terms of simulation in a simulation which is computationally a very inefficient procedure.

The main idea in the CDS option pricing using the fundamental PDE is to formulate the CDS quote K_τ for a tenor τ , which is the underlying CDS quote of the CDS option, as some unique tenor-specific function f_τ of the stochastic intensity,⁸⁷ that is $K_\tau = f_\tau(\gamma)$. This idea is perfectly consistent with the CDS pricing Section III.2.1.3 where the PDE (III.2.10) describes the *value* of the CDS contract with respect to the only stochastic variable γ . The only minor technical difference is that the PDE (III.2.10) was defined in terms of the absolute maturity T_i and in the world of the CDS options the underlying CDS is rather labelled with some tenor τ . In turn the 'function' f_τ is given implicitly by the whole PDE (III.2.10) (see the end of the Section III.2.1.3 where we have already opened this topic). Hence we conclude that the PDE (III.2.10), implicitly representing f_τ , allows us to find *either* (i) γ for which the value of the newly entered CDS contract is zero given CDS premium K_τ , *or* (ii) the CDS premium K_τ of a newly opened CDS contract for which the value of the CDS is zero given the intensity γ . While the former property was used in the estimation of parameters, the latter property is used in the CDS pricing and also in pricing of the CDS options.

We now give a more exact description of the CDS option pricing problem. Let $K_\tau(t, r, \gamma)$ denote the fair CDS spread related to a CDS with tenor τ at time t , given $r = r(t)$ and intensity $\gamma = \gamma(t)$ are stochastic variables. Hence $K_\tau(t, r, \gamma)$ is a random variable, yet specified as a function of both r, γ . The underlying CDS whose market quote is $K_\tau(t, r, \gamma)$ always covers the period $[t, t + \tau]$ and 'never matures'.⁸⁸ We define CDS call option $V(t, r, \gamma)$ on the CDS spread K_τ as a contract maturing at T when it provides its holder with the payoff

$$V(T, r(T), \gamma(T)) = g(r(T), \gamma(T)) = \max[K_\tau(T, r(T), \gamma(T)) - K, 0],$$

whose present value is in view of the standard risk-neutral valuation formula

$$V(t, r(t), \gamma(t)) = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T r(u) du \right) g(r(T), \gamma(T)) \middle| \mathcal{F}(t) \right]. \quad (\text{III.2.12})$$

To relax the negligible role of the interest rates we fix $r(t) = r$ to some constant and define $\tilde{K}_\tau(t, x_2) = K_\tau(t, \gamma) \equiv K_\tau(t, r, \gamma)$ and $\tilde{V}(t, x_2) = V(t, \gamma) = V(t, r, \gamma)$ to make it

⁸⁷we assume the risk-free rate is constant here. Should this assumption be relaxed then r would also have to be included in the function f_τ as an argument.

⁸⁸this means that e.g. the 5Y CDS quote will always be traded, despite the CDS tenor '5Y' denoting the life of the CDS contract.

clear r is not a stochastic variable anymore. It is easy to see by comparing (III.2.12) and (III.2.1) that we can immediately call Feynman-Kac to recover the pricing PDE for such an instrument $V(t, \gamma) \equiv V(t, r, \gamma)$ subject to the only Itô's process γ . The only but serious difficulty in (III.2.12) is the terminal payoff function g here because the fair CDS spread $\tilde{K}_\tau(t, x_2) = K_\tau(t, \gamma)$ has no closed-form with respect to γ (see the end of the Section III.2.1.3). On the other hand, we have already established a pricing relationship (III.2.10) between the CDS spread K_i and the value of the CDS contract \tilde{V}_i in terms of γ (or $x_2 = f(\gamma)$, equivalently). As a result, if we know that the value of the CDS contract $\tilde{V}_i(t)$ which matures at $T_i = t + \tau_i$ must be 0 at time t for some K_i given $x_2 = f(\gamma)$, then we can reverse this idea to conclude that there always exists some critical $K_i = K_i^*$ for a given γ which makes the CDS contract $\tilde{V}_i(t)$ zero-valued at t . This is the key idea in the identification of the function g . Denoting $x_2 = \ln \gamma$ as usual, the CDS option $\tilde{V}(t, x_2) = V(t, \gamma)$ satisfies

$$\begin{aligned} \frac{\partial \tilde{V}(t, x_2)}{\partial t} + \mathcal{A}\tilde{V}(t, x_2) - r\tilde{V}(t, x_2) &= 0 \\ \tilde{V}(T, x_2) &= \max[\tilde{K}_\tau^*(T, T + \tau, x_2) - K, 0], \end{aligned} \quad (\text{III.2.13})$$

with \mathcal{A} being the same operator as in (III.2.10) and $\tilde{K}_\tau^*(T, T + \tau, x_2)$ being such a value of $K_i = K_i^*$ in (III.2.10) for which $\tilde{V}_i(0, x_2) = 0$ (maturing at $T_i = \tau$). The pricing of credit spread options therefore requires at the first place a numerical search for $\tilde{K}_\tau^*(T, T + \tau, x_2^j)$, $j = 1, \dots, m$, where m is the number of spatial points in the x_2 -direction on the CDS option grid. This involves m root searches applied furthermore to the pricing function which calculates $\tilde{V}_i(0, x_2^j)$ according to the PDE (III.2.10). Under this setup, we have totally eliminated any possible complications which would be associated with modeling the dynamics of the credit spread itself which means we stick with the idea that the CDS spread at any time t is completely explained by the intensity process γ (and its parameters) which determines the default probability distribution of the reference name. As both the CIR and Black-Karasinski have shown to explain the CDS term structure almost perfectly (see the negligible errors in the Table 4 and Table 5), this is an empirically consistent assumption.

To demonstrate the pricing of CDS options on a real-world example, let us select a 5Y CDS ($\tau = 5$) written on Goldman Sachs as the underlying contract of the CDS option. Assume the CDS option matures in $T = 2$ years.⁸⁹ Also assume that the strike of the CDS option is $K = 0.01$ (hence we could say that the option is 'ITM' with respect to the current CDS quote of 115 bps.). Solving the PDE (III.2.13) the price of such a call CDS option has turned out to be 0.0063, or 63 bps in CIR whose corresponding pricing surface is shown in the Figure 14. The same task solved in terms of the calibrated Black-Karasinski gives the option premium 0.0069 (69 bps) and the pricing surface is presented in the Figure 15. We conclude that although both values do not coincide they are arguably very good estimates of the fair value of the CDS option especially when we do not take into account the historical behaviour of the CDS quotes. To capture the historical behaviour, we propose a different calibration approach.

⁸⁹the maturity of the option can be virtually any positive number. The idea is that the underlying CDS contract is always traded, regardless of its tenor τ .

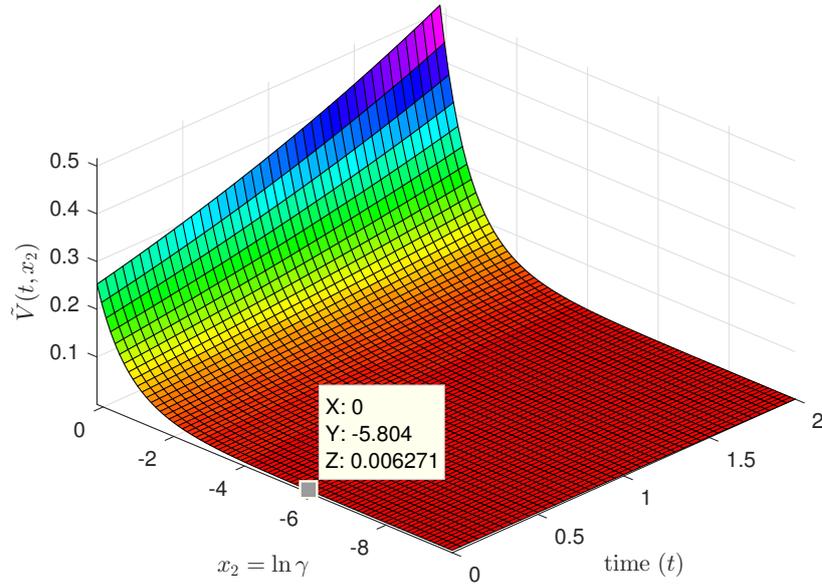


Figure 14: Log CIR pricing surface of 2Y call option with strike $K = 0.01$ on 5Y CDS written on Goldman Sachs. The pricing is as of March 25, 2016. The solution $\tilde{V}(0, x_2) = 0.0063$ lies at $x_2 = \ln(\gamma) = \ln(0.0029) = -5.8430$ (see the default intensity γ of Goldman Sachs in the Table 4).

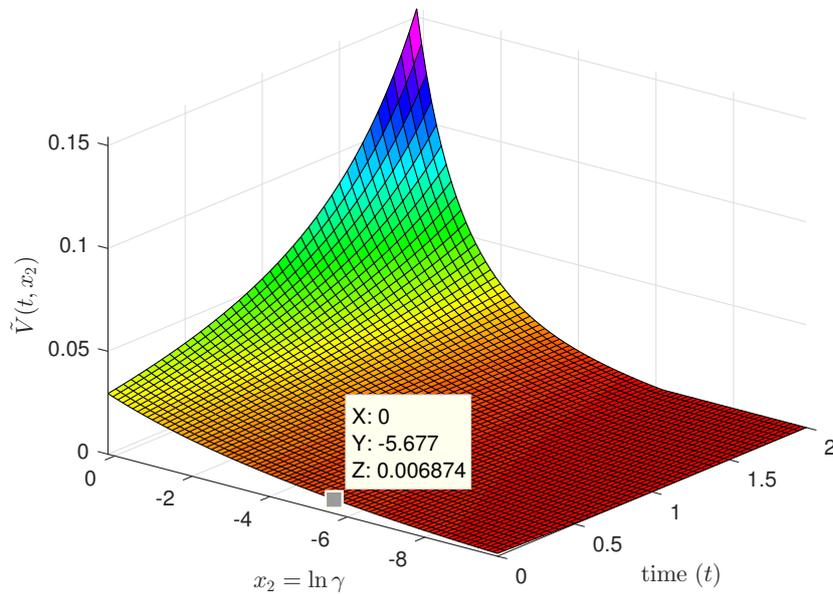


Figure 15: Black-Karasinski pricing surface of 2Y call option with strike $K = 0.01$ on 5Y CDS written on Goldman Sachs. The pricing is as of March 25, 2016. The solution $\tilde{V}(0, x_2) = 0.0069$ lies at $x_2 = \ln(\gamma) = \ln(0.0034) = -5.6740$ (see the default intensity γ of Goldman Sachs in the Table 5).

When pricing the CDS options it might not be optimal to use the parameters estimated from the 'current CDS curve'. Pricing of CDS options in the stochastic intensity framework is conceptually similar to the pricing of the interest-rate options using short-rate models. There, one is often confronted with the fact that even when the current yield curve is perfectly matched by the model, the prices of the interest-rate options populated by the model are not consistent with the prices quoted on the market.

The best approach in our case would, clearly, be to obtain the model parameters by fitting the prices of CDS options directly. These quotes are unfortunately not available as the product is rather exotic. On the other hand, we observe some time series of CDS quotes from which the model parameters could be estimated. We lay out this idea in the following way. If two CDS series (for example 5Y and 3Y quote) are available on a daily basis for a certain time horizon, then we assume that the parameters $\kappa_2, \theta_2, \sigma_2$ that determine the prices of the 3Y and 5Y CDS are constant through the whole period but γ changes every day. This is consistent with the CDS option pricing, where it is assumed that the only stochastic variable in the system is the intensity γ and this variable can explain the whole CDS curve. The difficulty here (compared to r -example) is that the time series of the default intensity is not observable. Even if it were, then the estimation of parameters from directly observing γ would be conducted under the physical measure \mathbb{P} instead of \mathbb{Q} and so it would lead to wrong results anyway. What we suggest is the following approach.

Assume we have a set of data which contains M observations of 3Y CDS quotes and M observations of 5Y CDS from which we would like to infer the intensity model parameters. We also have some candidate set of parameters $\Theta_2 = (\kappa_2, \theta_2, \sigma_2)$ (with γ being excluded) and we would like to calibrate the model for pricing of options on the 5Y CDS. Holding the parameter set Θ_2 fixed, for every day $m = 1, \dots, M$ we numerically find a critical value $\gamma_m^*(\Theta_2)$ such that it makes the auxiliary 3Y CDS zero valued. After this is done, we use this series of parameter-conditional implied hazard rates $\{\gamma_m^*(\Theta_2)\}_{m=1}^M$ and the parameter set Θ_2 to value the 5Y deals. If the parameters Θ_2 are optimal then every 5Y CDS deal must have zero value. The mathematical formulation of this whole procedure is

$$\hat{\Theta}_2 = \arg \min_{\Theta_2} \left[\sum_{m=1}^M |V_m^{5YCDS}(\gamma_m^*(\Theta_2), \Theta_2)| \right]$$

$$\gamma_m^*(\Theta_2) = \{ \gamma_m : V_m^{3YCDS}(\gamma_m, \Theta_2) = 0 \}.$$

The idea of including the 'auxiliary' time series of 3Y CDS quotes is necessary to identify the latent variable $\gamma_m^*(\Theta_2)$, conditional on Θ_2 for every day m . If we did not include the second 3Y CDS time series then we would have M unknowns $\gamma_m^*(\Theta_2), m = 1, \dots, M$ plus 3 unknown parameters $(\kappa_2, \theta_2, \sigma_2)$ and only M observations of the time series 5Y CDS quotes. Clearly, this would be an undeterminate system with no unique solution. Also, see the mechanics of the optimization. If Θ_2 is far from being optimal, then to make the 3Y deals zero-valued $\gamma_m^*(\Theta_2)$ has to be 'wild' to compensate the imperfection of Θ_2 . In turn the sum $\sum_{m=1}^M |V_m^{5YCDS}(\gamma_m^*(\Theta_2), \Theta_2)|$ will most likely be 'wild' too which forces the optimization algorithm to find better Θ_2 . In the ideal case, when all 3Y and 5Y deals are historically

determined by the same set of parameters Θ_2 (and by $\gamma_m^*(\Theta_2), m = 1, \dots, M$), the sum is zero. We also can say that these parameters are *historically risk-neutral* because they have been obtained by matching quoted prices in a historical time series. This is probably the most accurate and meaningful approach of fitting the parameters when it can not be obtained directly by matching the option data. It captures the restricted term-structure represented by 3Y and 5Y CDS quotes here, yet it also takes into account the historical behaviour of the CDS series.

III.2.2 Extensions

III.2.2.1 Counterparty risk and extension for correlated defaults

In this part, we introduce a general method of extracting counterparty credit risk from a transaction with counterparty subject to default risk. The method is as general and as powerful as the fundamental pricing PDE (III.2.2) and so it can be useful in many applications. The main purpose of the previous sections was to investigate the pricing and valuation of contingent claims subject to default risk, and we now take these ideas one step further and allow for additional source of credit risk to enter the pricing and valuation. This additional source of credit risk can be substantial especially when the contingent claim itself is written on credit risk. A good example can be a CDS option whose value and payoff depends on the credit risk (induced by some intensity) of the reference name. It is well known that the credit market participants are typically not completely independent entities and thus there could be a substantial *counterparty credit risk* in a transaction of this kind. In particular, if the payoff of the CDS option is e.g. $\max[\text{CDS quote} - K, 0]$ then the so called wrong way risk may play a major role in the fair price of such a transaction. The wrong way risk factor is quite obvious here. If the CDS spread is high then the credit of the reference entity has worsened (and the potential payoff increases) but as our counterparty might not be independent from the reference entity, its probability of default could increase too.

The main idea of calculating the counterparty credit risk is to define a variable representing the loss due to counterparty's default and re-iterate the previous results. We show that even this random loss satisfies some particular PDE and thus the credit risk can be easily computed by the standard PDE methods. We also show that it is possible to introduce a correlation of defaults through correlated intensity processes and demonstrate its key role in the wrong way risk and the good way risk. We start with a discussion on default correlation in intensity models and then we introduce the counterparty credit risk PDE.

In the intensity framework the incorporation of 'correlated defaults' is less straightforward than in the firm-value models. In the firm-value framework, one typically assumes that the asset value processes are correlated and this immediately generates sample paths of firm asset values that more or less move together and possibly hit some default barriers.⁹⁰ In other words, if the asset value processes are perfectly correlated and the asset value of one company decline for a certain period it will default and so will the other company (considering both companies have the same default threshold). In the intensity

⁹⁰see the *First passage time* models in the Section II.1.2.

framework the default is, however, always a random event regardless of the value of the intensity. This is the basic property of total inaccessibility of the random default time τ^* . If we assume two distinct names $i \neq j$ such that $\gamma_i(t) = \gamma_j(t)$ for all t , then still their defaults will not be perfectly correlated and likely the default correlation will be low. In turn, regardless of the intensity, we can not certainly say that both companies default in a given time-frame even if their default intensities are perfectly correlated processes. For further details on the default correlations we refer the reader to (Duffie, 2003, Section 10).

Let us discuss the default correlation in the intensity-based models first. Assume as usual $H_i(t) = \mathbf{1}_{\{\tau_i^* \leq t\}}$ being the default indicator of the name i . The correlation $\rho_{i,j}(t)$ for names $i \neq j$ expresses the correlation of H_i, H_j on the time interval $[0, t]$ and it equals

$$\begin{aligned} \rho_{i,j}(t) &= \text{corr}[H_i(t), H_j(t)] \\ &= \frac{\mathbb{E}[H_i(t)|\mathcal{F}(0)] + \mathbb{E}[H_j(t)|\mathcal{F}(0)] - \mathbb{E}[H_{i,j}(t)|\mathcal{F}(0)] - \mathbb{E}[H_i(t)|\mathcal{F}(0)]\mathbb{E}[H_j(t)|\mathcal{F}(0)]}{\sqrt{\mathbb{E}[H_i(t)|\mathcal{F}(0)](1 - \mathbb{E}[H_i(t)|\mathcal{F}(0)])}\sqrt{\mathbb{E}[H_j(t)|\mathcal{F}(0)](1 - \mathbb{E}[H_j(t)|\mathcal{F}(0)])}}, \end{aligned} \quad (\text{III.2.14})$$

where

$$H_{i,j}(t) = \mathbf{1}_{\{\tau_{i,j}^* \leq t\}}, \text{ with } \tau_{i,j}^* = \min[\tau_i^*, \tau_j^*],$$

is the indicator of 'i or j' being in default. The intensity of either 'i or j' being in default is then addressed by the following Lemma.

Lemma III.2.3 (Intensity representation of the default time $\tau_{i,j}^*$)

Assume $i \neq j$ and $\mathbb{P}[\tau_i^* = \tau_j^*] = 0$. Then

$$\gamma_{i,j} = \gamma_i + \gamma_j,$$

is the intensity process for

$$\tau_{i,j}^* = \min[\tau_i^*, \tau_j^*].$$

The moral of the Lemma III.2.3 is that we can easily compute the probability

$$\mathbb{P}[\tau_{i,j}^* \leq t | \mathcal{F}(0)] = \mathbb{E}[H_{i,j}(t) | \mathcal{F}(0)] = 1 - \mathbb{E}\left[\exp\left(-\int_0^t (\gamma_i(u) + \gamma_j(u))du\right) \middle| \mathcal{F}(0)\right],$$

which is a critical component of the default correlation $\rho_{i,j}(t)$ in (III.2.14). It is also easy to see that if γ_i, γ_j are *independent* processes then

$$\begin{aligned} \mathbb{P}[\tau_{i,j}^* \leq t | \mathcal{F}(0)] &= 1 - \mathbb{E}\left[\exp\left(-\int_0^t \gamma_i(u)du\right) \middle| \mathcal{F}(0)\right] \mathbb{E}\left[\exp\left(-\int_0^t \gamma_j(u)du\right) \middle| \mathcal{F}(0)\right] \\ &= 1 - \mathbb{P}[\tau_i^* > t | \mathcal{F}(0)] \mathbb{P}[\tau_j^* > t | \mathcal{F}(0)] \\ &= \mathbb{P}[\tau_i^* \leq t | \mathcal{F}(0)] + \mathbb{P}[\tau_j^* \leq t | \mathcal{F}(0)] - \mathbb{P}[\tau_i^* \leq t | \mathcal{F}(0)] \mathbb{P}[\tau_j^* \leq t | \mathcal{F}(0)]. \end{aligned}$$

If we then substitute this 'independence' result into the formula for the correlation of default indicators $\rho_{i,j}(t)$ in (III.2.14), then immediately $\rho_{i,j}(t) = 0$ for any t . Therefore,

we see that the component $\mathbb{E}[H_{i,j}(t)|\mathcal{F}(0)]$ plays a major role in the default correlation $\rho_{i,j}(t)$.

The inability of the default intensities to model (highly) correlated defaults is a well-known issue and has been documented in many studies. In particular, (Hull and White, 2001) first object that even perfectly correlated intensities do not yield high default correlations. (Schönbucher and Schubert, 2001) come with the finding that the range of default correlations achievable by the reduced-form models is quite limited, below the empirically observed correlations and so the authors try to circumvent this by copula-based modeling and introduction of intensities that increase after the related entity's default. This sudden change of the default intensity after the default of other entity is also addressed by (Yu, 2007) who calls such an intensity specification the *total hazard*. (Cerny and Witzany, 2014) apply a completely different approach and they model correlation as the dependence between the *levels* of processes. For further discussion of the correlation modeling pitfalls in the reduced-form models we recommend reading (Glasserman, 2003, Section 9.4.2.) and (Schmid, 2004, Section 4.3)).

Having discussed the correlation of defaults, we now build up the framework for counterparty credit risk allowing for the correlated defaults. Assume being a default-free market participant and a further assume a deal, possibly written on credit risk of some reference entity, which has been entered into with a counterparty subject to credit risk. Focusing on the default event of the counterparty $\{\tau_\delta^* \leq T\}$ prior or at the maturity T of the deal, a recovery claim Z_δ is settled exactly at the counterparty's default time τ_δ^* . Denoting V_{rf} the hypothetical default-free value of the deal, let us consider the following scenarios:

- If $\tau_\delta^* \leq T$ and $V_{\text{rf}}(\tau_\delta^*, T) < 0$, then $Z_\delta(\tau_\delta^*) = V_{\text{rf}}(\tau_\delta^*, T)$. That is, after the counterparty's default we have to settle the full (risk-free) value with the defaulted counterparty, i.e. we *pay* our *obligations* amounting $V_{\text{rf}}(\tau_\delta^*, T)$ *in full*.
- If $\tau_\delta^* \leq T$ and $V_{\text{rf}}(\tau_\delta^*, T) > 0$, only a fraction $0 \leq \delta < 1$ of the positive exposure $V_{\text{rf}}(\tau_\delta^*, T) > 0$ is settled and so $Z_\delta(\tau_\delta^*) = \delta V_{\text{rf}}(\tau_\delta^*, T)$. That is, we *receive* only a *fraction* δ of the risk-free value of the deal $V_{\text{rf}}(\tau_\delta^*, T)$.

We can combine both scenarios into one by setting

$$Z_\delta(\tau_\delta^*) = V_{\text{rf}}(\tau_\delta^*, T)\mathbf{1}_{\{V_{\text{rf}}(\tau_\delta^*, T) < 0\}} + \delta V_{\text{rf}}(\tau_\delta^*, T)\mathbf{1}_{\{V_{\text{rf}}(\tau_\delta^*, T) > 0\}},$$

as the recovery payment (either positive or negative) just after the counterparty's default. Assume that we have entered a deal with a risk-free counterparty in the sense that it always pays its obligations in full even if it defaulted. Then the after-default settlement equals

$$Z_\delta(\tau_\delta^*)_{\delta=1} = V_{\text{rf}}(\tau_\delta^*, T) = V_{\text{rf}}(\tau_\delta^*, T)\mathbf{1}_{\{V_{\text{rf}}(\tau_\delta^*, T) < 0\}} + V_{\text{rf}}(\tau_\delta^*, T)\mathbf{1}_{\{V_{\text{rf}}(\tau_\delta^*, T) > 0\}}.$$

If we set $L_\delta(\tau_\delta^*) > 0$ as the *loss in market value* of the default-free deal V_{rf} due to counterparty's default and incomplete recovery we find

$$L_\delta(\tau_\delta^*) = Z_\delta(\tau_\delta^*)_{\delta=1} - Z_\delta(\tau_\delta^*) = (1 - \delta)V_{\text{rf}}(\tau_\delta^*, T)\mathbf{1}_{\{V_{\text{rf}}(\tau_\delta^*, T) > 0\}},$$

which is a *positive* quantity representing incurred credit loss caused by the counterparty's default. Henceforth we assume more specifically that $L_\delta(t) = L_\delta(t, x_1, x_2, x_\delta)$ is a function of time t , stochastic risk-free rate $x_1 = r$, default intensity of the reference credit x_2 and the default intensity x_δ of the deal counterparty. Since the loss is incurred only in the case of default prior or at the maturity T (otherwise there is no loss), the probability distribution of the counterparty's default time τ_δ^* has to be accounted for. We are now looking for the present value of this loss amount. Since it is incurred just after the counterparty's default, the discounting factor only takes into account the period from t until τ_δ^* and so the risk-neutral present value of this loss amount is

$$V_\delta(t, T) = \mathbf{1}_{\{\tau_\delta^* > t\}} \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{\tau_\delta^*} x_1(u) du \right) L_\delta(\tau_\delta^*, x_1(\tau_\delta^*), x_2(\tau_\delta^*), x_\delta(\tau_\delta^*)) \middle| \mathcal{F}(t) \right]. \quad (\text{III.2.15})$$

$V_\delta(t, T)$ represents the present value of the expected loss in the deal caused by the counterparty's default, in other words it is the CVA of the deal. Since in (III.2.15) we have Itô's processes x_1, x_2, x_δ , the expectation is solvable using the Feynman-Kac theorem. We in turn have the following theorem.

Theorem III.2.5 (Fundamental CVA equation for claims written on credit)

Let $x_1 = r$ be a stochastic risk-free rate and $x_2 = \gamma$ be the default intensity of the reference name. If a deal V written on the reference name, maturing at T has been entered into with a risky counterparty with default intensity $x_\delta = \gamma_\delta$, then the CVA of the contract, denoted by $V_\delta(t, x_1, x_2, x_\delta) = V_\delta(t, x)$ is on the set $\{\tau_\delta^* > t\}$ a solution to

$$\begin{aligned} \frac{\partial V_\delta(t, x)}{\partial t} + \mathcal{A}V_\delta(t, x) + x_\delta L_\delta(t, x) - (x_1 + x_\delta)V_\delta(t, x) &= 0 \\ V_\delta(T, x) &= 0, \end{aligned} \quad (\text{III.2.16})$$

where

$$\begin{aligned} \mathcal{A} &= \mu_1(t, x) \frac{\partial}{\partial x_1} + \mu_2(t, x) \frac{\partial}{\partial x_2} + \mu_\delta(t, x) \frac{\partial}{\partial x_\delta} \\ &+ \frac{1}{2} \sigma_1^2(t, x) \frac{\partial^2}{\partial x_1^2} + \frac{1}{2} \sigma_2^2(t, x) \frac{\partial^2}{\partial x_2^2} + \frac{1}{2} \sigma_\delta^2(t, x) \frac{\partial^2}{\partial x_\delta^2} \\ &+ \rho_{12}(t, x) \sigma_1(t, x) \sigma_2(t, x) \frac{\partial^2}{\partial x_1 \partial x_2} + \rho_{1\delta}(t, x) \sigma_1(t, x) \sigma_\delta(t, x) \frac{\partial^2}{\partial x_1 \partial x_\delta} \\ &+ \rho_{2\delta}(t, x) \sigma_2(t, x) \sigma_\delta(t, x) \frac{\partial^2}{\partial x_2 \partial x_\delta}, \end{aligned}$$

and where furthermore

$$L_\delta(t, x) = (1 - \delta) V_{\text{rf}}(t, x_1, x_2) \mathbf{1}_{\{V_{\text{rf}}(t, x_1, x_2) > 0\}}, 0 \leq \delta < 1,$$

is the credit loss incurred from the default of the counterparty in a deal whose fair value is $V_{\text{rf}}(t, x_1, x_2)$ at the counterparty's time of default $t = \tau_\delta^*$. The fair, counterparty credit risk adjusted value of the deal is

$$V(t, T, x_1, x_2, x_\delta) = V_{\text{rf}}(t, x_1, x_2) - V_\delta(t, x_1, x_2, x_\delta).$$

Remarks. The value of $V_{\text{rf}}(t, x_1, x_2)$ can either have a closed-form in terms of the spatial variables x_1, x_2 but more often it is a solution to the PDE (III.2.2) where V is replaced by V_{rf} .

For the practical applications it is, however, better to transform the variables as necessary (e.g. the log transform of the intensities is a common technique) to maintain a better stability (convergence)⁹¹ of the finite difference scheme. Also, in many cases it is useful to reduce the complexity of the scheme by one factor by fixing the factor that has the smallest contribution to the price. We provide two examples of the application of the CVA PDE.

Example 1: CDS contracts with defaultable counterparty. For illustration of the CVA equation (III.2.16), we analyze two uncollateralized deals with Credit Suisse (CS) which have been entered into some time ago, both are CDS written on Bank of America (BoA) and mature in 6 years. We analyze the first deal completely separately from the other one. In the first CDS deal we act as a protection seller (CS is a protection buyer) and the deal runs under $K = 0.04$. In the second deal, we are protection buyer (CS is a protection seller) and the deal runs under $K = 0.005$. We have selected these values of K intentionally, so that both positions (protection seller (PS) and protection buyer (PB)) are 'in the money' in order that the CVA is more visible. We further assume that intensities of CS and BoA are either both driven by CIR (parameters in the Table 4) or BK (Table 5). After solving the PDE III.2.16 the CVA profiles of all these combinations with respect to the 'intensities' correlation parameter ρ is shown in the Figure 16.

The Figure 16 reveals a lot of useful information. Firstly, we can see that CVA of the PS position is a decreasing function of ρ while the CVA of PB position is an increasing function of ρ . As we would in reality expect the intensities to be positively correlated, PS in such case experiences GWR while the BP experiences a WWR. This is a general result for $\rho > 0$. Secondly, it is easy to see that the model specification has a significant effect on the ability to capture the correlation. Obviously the outputs from the BK model are much more reactive to the value of ρ while the CVA computed using CIR reacts to ρ only modestly. The third observation is that the CVA around $\rho = 0$ is nearly the same for the CIR and the BK process. This was expected because it would be strange that two models (fitted to the same data) give significantly different results while the independence of the intensities is considered.

The CVA PDE (III.2.16) can furthermore be used to identify some special features of the CVA. We in particular found that the usual assumption that riskier counterparty relative to a safe counterparty (*ceteris paribus*) implies a higher CVA does not hold absolutely. The Figure 17 illustrates this. The main finding in this CVA profile is that the CVA increases with increasing counterparty's (log)intensity $\ln(\gamma_\delta)$ yet only up to a certain level of the intensity where interestingly the CVA starts to decrease. Hence,

⁹¹rigorously, the relationship between the stability and convergence is addressed by the *Lax Equivalence Theorem* which states: consistency + stability \Leftrightarrow convergence of a finite difference scheme. See (Crépey, 2013, p. 217).

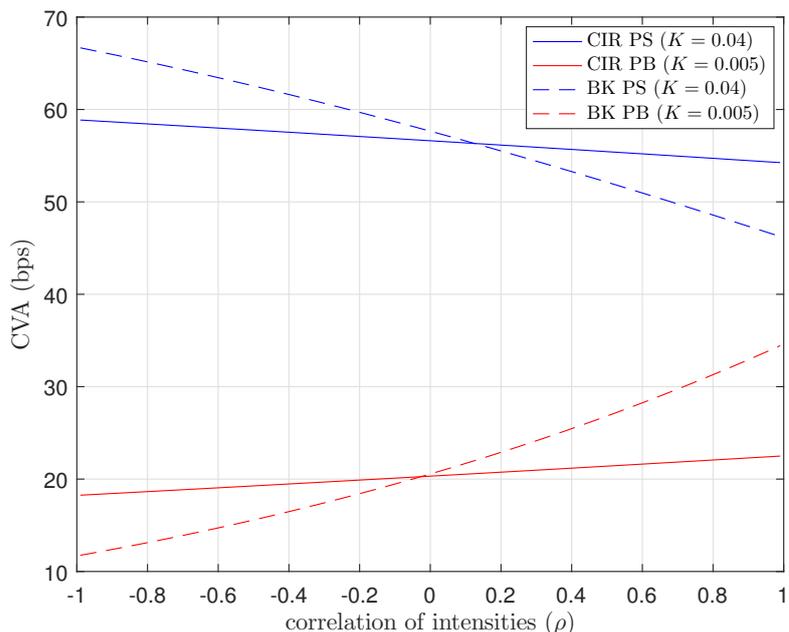


Figure 16: CVA as a function of correlation of the default intensities ρ . The deal is a CDS with the counterparty Credit Suisse and written on the credit of Bank of America. The deals are assumed to have 6 years to maturity and two cases of K are considered. $K = 0.04$ for the protection seller deal and $K = 0.005$ for the protection buyer deal.

some sort of a 'hump' shape of the CVA originates and this CVA hump region is well visible in the Figure 17. In turn we can conclude that a deal with very (up to extremely) risky counterparty might eventually be less risky than the same deal with moderately risky counterparty. We also found out after doing several numerical experiments that the hump is typically more noticeable when the CDS deal is deeply 'out-of-the money'. This, however, also can be justified by the fact that the out-of-the money CDS deal can get risky (in terms of the credit risk) only when the exposure increases.⁹² With a very risky counterparty, defaulting probably very soon, we thus have nothing to lose given OTM money deal. When, however, the counterparty were less risky, and the deal were OTM at the same time, then it could very well happen that the exposure increases *before* the counterparty defaults and hence there is a potential counterparty credit risk. This is exactly captured by the CVA hump. On the other hand, when we consider a very ITM deal, then the exposure can not go much further (note the intensity of the underlying CDS entity is driven by a mean-reverting process) and will likely decline in the future. Hence, soon counterparty's default generates a large credit loss and in turn the riskier counterparty, the higher CVA and there is no CVA hump.

The Table 6 summarizes all the possible scenarios and effects of the intensity correlation ρ on the counterparty credit risk in a CDS deal from the perspective of the protection seller

⁹²when a counterparty defaults given MtM is negative, this does not count as a *credit* loss.

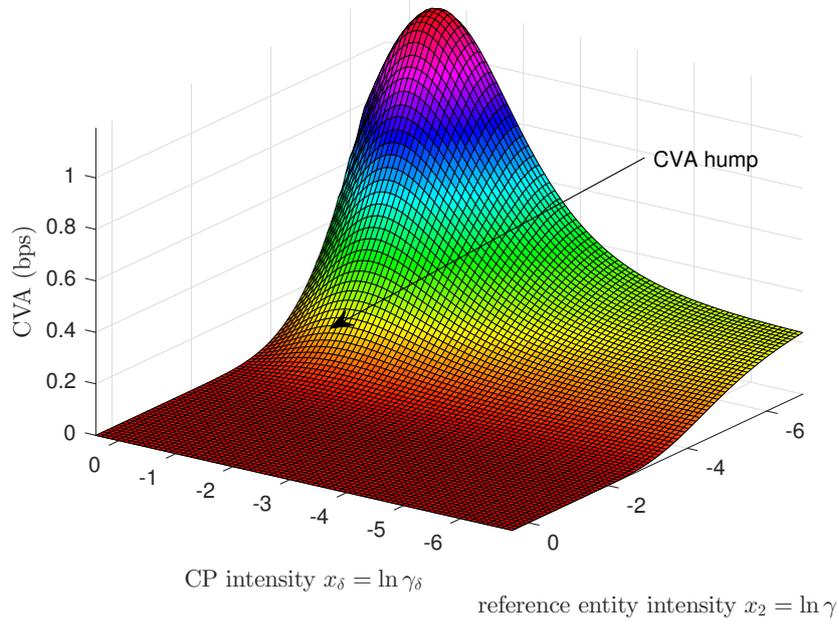


Figure 17: CVA surface in bps of a CDS deal written on BoA, entered into with CS as a protection seller at $K = 0.01$. The deal matures in 6 years, BoA and CS default intensities are assumed to be correlated ($\rho = 0.8$) and both intensities follow the CIR dynamics. Notice the CVA hump region.

and buyer. It is easy to see from the table that ρ has an opposite effect on the position of the PB and PS. Since ρ is, in general, positive the PS has *ceteris paribus* an advantage over the PB because the PS experiences GWR while the PB experiences WWR when the deal has a positive value.

From this detailed example it is easy to see how much information the fundamental CVA equation (III.2.15) can provide about the credit risk in the transaction. The equation, however, does not necessarily need be applied to contingent claims only written on credit risk. It can be applied to assess the *counterparty* credit risk of any transaction whose pricing is subject to Itô's processes, yet that is always just an analogy of the case we have investigated. We demonstrate this on a simple Black-Scholes example.

Example 2: Stock option with defaultable counterparty. Assume a long call stock option position written on Bank of America (BoA) and that it is March 25, 2016. The option matures on Jan 19, 2018 and its strike is USD 13. Our counterparty in the deal is Credit Suisse (CS) whose CIR intensity has the parameters from the Table 4. It is assumed that when CS defaults, we recover $\delta = 0.4$ of the outstanding exposure. The stock price (BoA) at the time of valuation is USD 13.68 and the option trades at USD 2.53. We calculate the implied volatility of the stock option to be $\sigma = 0.2928$. Having the inputs it can be shown that the CVA PDE (III.2.16) can be slightly modified in order to calculate the CVA of the deal. Solving the PDE (III.2.16) to calculate the CVA, we have the CVA $V_\delta = \text{USD } 0.0528$, or 209 bps of the deal's risk-free value if we assume no

$V_{\text{rf}}(t, T)$	ρ	CP's intensity γ_δ		credit risk		effect	
		PS	PB	PS	PB	PS	PB
+	+	low	high	low	high	GWR	WWR
+	-	high	low	high	low	WWR	GWR
-	+	low	high	low	low	-	-
-	-	high	low	low	low	-	-

Table 6: Analysis of the effect of default intensity correlation ρ in a CDS deal. $V_{\text{rf}}(t, T)$ represents the risk-free value of the CDS position. ρ denotes the correlation of the counterparty's default intensity γ_δ and intensity of the reference name γ . *CP's intensity* shows what level of intensity is the counterparty likely to have according to $V_{\text{rf}}(t, T)$ and ρ .

correlation ρ between the diffusion of the CIR intensity and the stock price process. If we consider the highly unlikely case of $\rho = 1$, then the CVA *increases* to USD 0.0658 and when $\rho = -1$ then the CVA is only 0.0412 USD. Thus, the correlation ρ has a significant impact on the CVA. It is, however, natural to assume that $\rho < 0$ due to the fact that if BoA's stock goes up, then it is likely that the stock of CS also goes up which means its default intensity actually decreases. Hence, there is, in general a good way risk in the case of $\rho < 0$ for the long call position. The opposite, however, holds for long put option position, where $\rho < 0$ implies wrong way risk. There, if BoA slumps and the option gets more valuable, the stock of CS is also likely to slump which means its intensity goes up and hence greater probability of default when the exposure grows.

III.2.2.2 Multidimensional jump–diffusion model and PIDE

In the previous sections, we have illustrated how the pricing, calibration and computations are done under intensities that are Itô's processes (following the dynamics of the class (I.1.1)). In this section, we extend this to jump-diffusions with Poisson jumps and present a Common Poisson Shock (CPS) model for a more elaborate modeling of correlated defaults in the scope of the reduced-form framework. We show that while the intensity subject to Itô's process results into a specific PDE, the intensity subject to jump-diffusion results into an analogous partial integro-differential equation (PIDE). The main motivation for the introduction of jumps stems from the empirical observations. The Figure 18 shows evolution of CDS quotes of several big names (+ on-the-run iTraxx index), where both idiosyncratic and some sort of 'systemic' jumps in CDS quotes are easy to be identified (see (Cariboni and Schoutens, 2009) for a list of examples). Although the jumps in CDS prices are not the same quantities as the jumps in intensity, both these credit quantities are directly related.⁹³ If we are about to prove the presence of the jumps in the intensities mathematically/statistically, we need to define a test of the relevance of the jump component in the dynamics of the intensity. For this purpose, we have selected the

⁹³this can best be illustrated using the 'credit triangle' which connects intensity, recovery and credit spread through a simple, yet surprisingly applicable relationship: CDS spread = $\gamma(1 - \delta)$.

following dynamics to describe the evolution of the log-intensities ($x = \ln \gamma$)

$$dx(t) = \kappa(\theta - x(t))dt + \sigma dW(t) + J(t)dN(t), \text{ with } J(t) \sim U[a, b], \quad (\text{III.2.17})$$

where U stands for the uniform distribution. This essentially mimics the log-stock specification from (Synowiec, 2008) yet the drift is different. The reason for this choice of the jump component is simply that it allows for the user-controlled separation of the jumps and the diffusion because only sudden changes in x of at least some amplitude (subject to the parameters $0 < a < b$) classify as jumps. After analyzing the data,⁹⁴ we have resorted to defining the jump as a sudden change of $x = \ln \gamma$ of at least $a = 0.3$ to filter out the diffusion. The results from the estimation of the parameters are shown in the Table 7 where it is easy to see that the intensity of jumps λ implies a jump of amplitude at least $a = 0.3$ every 2 up to 3 years, and that all the four credits exhibit relatively similar jump behaviour. It thus makes sense to develop models where the intensity is a jump-diffusion process.

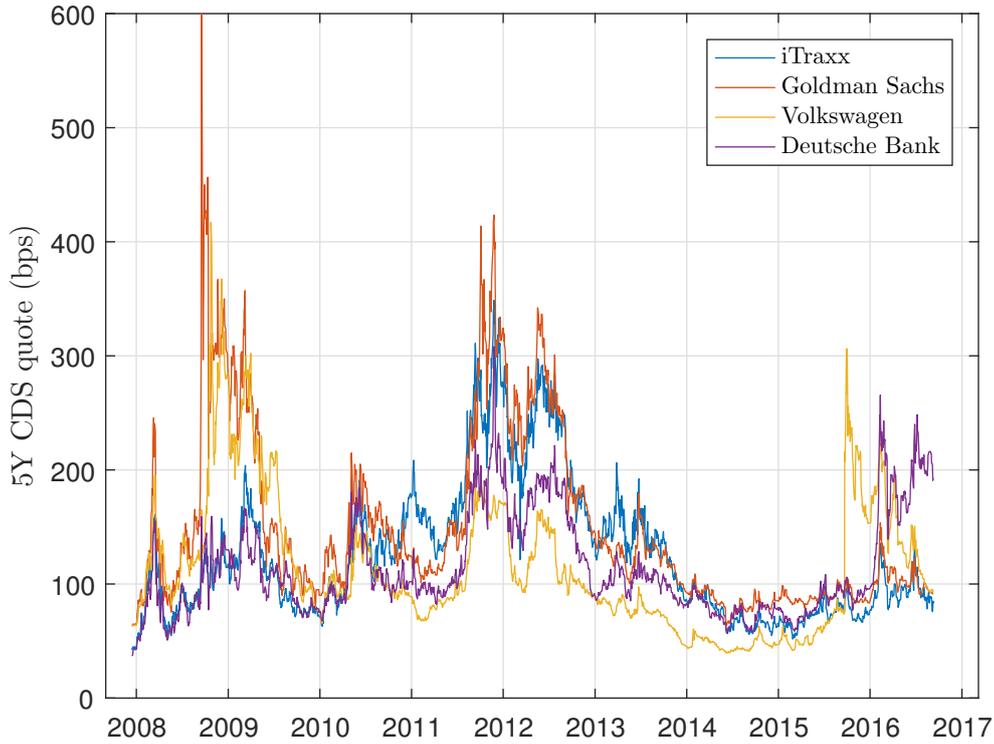


Figure 18: The last decade evolution of 5Y CDS quotes of selected credits.

Let us now provide a mathematical formulation to a problem with intensity subject to a jump-diffusion. Since the only two conditions for the validity the Theorem III.2.1 are that all processes x_i are Markovian and the expectation is finite (this also involves a sufficient integrability) the result is also valid for Markovian jump-diffusions with the only difference

⁹⁴that is the logarithms of γ .

credit/parameter	κ	θ	σ	λ
iTraxx	1.9586	-3.9981	0.7818	0.3964
GS	2.0555	-3.8863	0.7009	0.4961
VW	1.9127	-4.0121	0.6730	0.4624
DB	1.3084	-4.2503	0.5409	0.5329

Table 7: Estimated parameters of the jump-diffusion equation (III.2.17) for log-intensities on a last decade time series data for various names.

in the resulting equation which is now PIDE. Finding a pricing equation then only involves finding an appropriate form of the infinitesimal generator \mathcal{A} which generates the f -process such that $f(t, x) = \mathbb{E}[\phi(x(T)) | \mathcal{F}(t)]$. Because

$$\mathcal{A}f(t, x) = \lim_{\Delta t \rightarrow 0} \left(\frac{\mathbb{E}[f(t + \Delta t, x) | \mathcal{F}(t)] - f(t, x)}{\Delta t} \right) = \frac{\mathbb{E}[df(t, x) | \mathcal{F}(t)]}{dt},$$

then Itô's lemma for the jump processes implies

$$\mathcal{A} = \mathcal{A}_c + \mathcal{A}_J,$$

where \mathcal{A}_c is the generator of the diffusion and \mathcal{A}_J is the pure jump generator (for details see also (Gikhman, 1972, p. 299)). When combined with (III.2.1) this directly leads to a Proposition.

Proposition III.2.1 (Pricing equation for jump diffusing intensity)

Assume a claim $V(t, f_1(x_1), f_2(x_2)) = \tilde{V}(t, x_1, x_2)$, where x_1 is a market variable process and x_2 is a process of counterparty's default intensity γ or any transformation of it. When the processes x_1, x_2 follow

$$\begin{aligned} dx_1(t) &= \mu_1(t, x(t))dt + \sigma_1(t, x(t))dW_{x_1}^{\mathbb{Q}}(t) \\ dx_2(t) &= \left(\mu_2(t, x(t)) - a_2\lambda(t)\mathbb{E}^{\mathbb{Q}}[J(t)] \right) dt + \sigma_2(t, x(t))dW_{x_2}^{\mathbb{Q}}(t) + a_2J(t)dN(t), \end{aligned}$$

where $dW_{x_1}^{\mathbb{Q}}(t)dW_{x_2}^{\mathbb{Q}}(t) = \rho(t, x_1, x_2)dt$, N is a Poisson process with intensity λ , J is the magnitude of the jump with PDF $f_{J(t)}$ and a_2 is a constant, then the claim \tilde{V} satisfies the PIDE

$$\frac{\partial \tilde{V}}{\partial t} + \mathcal{A}_c \tilde{V} + \mathcal{A}_J \tilde{V} + m(t, x_1, x_2) = 0.$$

Here \mathcal{A}_c is the infinitesimal generator of the diffusion components in x_1, x_2 and

$$\mathcal{A}_J \tilde{V} = \lambda(t) \int_D \left(\tilde{V}(t, x_1, x_2 + a_2 z) - \tilde{V}(t, x_1, x_2) \right) f_{J(t)}(z) dz.$$

The term m consists of the 'Lando' terms \tilde{Z}, \tilde{c} introduced earlier.

This approach of deriving the PIDE, however, can be reiterated further and we use it for the Common Poisson Shock (CPS) model.

As pointed out in the preceding Section III.2.2.1, the often discussed disadvantage of the reduced-form models is the typically low achievable default correlation although the correlation parameter of the intensities ρ is set to its technical maximum 1.⁹⁵ We show that inclusion of 'simultaneous' jumps can lead to significantly higher default correlation and that even such a complex specification results into a PIDE with a *single* one-dimensional convolution integral. This in turn does not dramatically increase the scheme's computational complexity. We call this specification the Common Poisson Shock model. We firstly give a Proposition which is then followed by a deeper discussion.

Proposition III.2.2 (The Common Poisson Shock and its PIDE)

Consider a claim $V(t, f_1(x_1), f_2(x_2)) = \tilde{V}(t, x_1, x_2)$, where either x_1 or x_2 is a process of counterparty's default intensity or any transformation of it. When the processes x_1, x_2 follow

$$\begin{aligned} dx_1(t) &= \left(\mu_1(t, x_1, x_2) - a_1 \lambda(t) \mathbb{E}^{\mathbb{Q}}[J(t)] \right) dt + \sigma_1(t, x_1, x_2) dW_{x_1}^{\mathbb{Q}}(t) + a_1 J(t) dN(t) \\ dx_2(t) &= \left(\mu_2(t, x_1, x_2) - a_2 \lambda(t) \mathbb{E}^{\mathbb{Q}}[J(t)] \right) dt + \sigma_2(t, x_1, x_2) dW_{x_2}^{\mathbb{Q}}(t) + a_2 J(t) dN(t), \end{aligned}$$

where $dW_{x_1}^{\mathbb{Q}}(t) dW_{x_2}^{\mathbb{Q}}(t) = \rho(t, x_1, x_2) dt$, then the infinitesimal generator \mathcal{A} can be decomposed into a purely continuous generator \mathcal{A}_c and pure jump generator \mathcal{A}_J given as

$$\mathcal{A}_J \tilde{V} = \lambda(t) \int_D \left(\tilde{V}(t, x_1 + a_1 z, x_2 + a_2 z) - \tilde{V}(t, x_1, x_2) \right) f_{J(t)}(z) dz,$$

and the claim \tilde{V} satisfies the PIDE

$$\frac{\partial \tilde{V}}{\partial t} + \mathcal{A}_c \tilde{V} + \mathcal{A}_J \tilde{V} + m(t, x_1, x_2) = 0, \quad (\text{III.2.18})$$

where m consists of the 'Lando' terms \tilde{Z}, \tilde{c} introduced earlier.

The main 'new' element in the jump-diffusion models, compared to the Itô-based models introduced earlier is the *pure* jump component $a_i J(t) dN(t)$. The $a_i \lambda(t) \mathbb{E}^{\mathbb{Q}}[J(t)]$ in the drifts is the compensator term.⁹⁶ The critical point here is that N is the *common* Poisson process (with nonrandom intensity $\lambda(t)$) dictating the simultaneous jumps in both x_1, x_2 . If a jump occurs at t , then each process after the jump is set to $x_i(t) = x_i(t-) + a_i J(t)$, where $J(t)$ is an independent random variable with density $f_{J(t)}$ and $a_i J(t)$ is the magnitude of the jump for the i th process. The convolution integral in the pure jump generator \mathcal{A}_J then captures the instantaneous change in the claim's value \tilde{V} as a result of a jump in both x_1, x_2 . Notice that we have specified the model such that only a *one-dimensional* integral appears. This is a desired property because if we used J_1, J_2 instead of J then we would have a two-dimensional integral instead which is a much bigger numerical obstacle in the numerical scheme. A specification of J and also a

⁹⁵also see the low impact of the Wiener processes correlation ρ on the CVA under CIR-intensity in the Figure 16.

⁹⁶to ensure that the compensated jump process is a \mathbb{Q} -martingale.

possible calibration of the dynamics is a delicate task.⁹⁷ We do not investigate this further and refer the reader to (Synowiec, 2008), who discusses Normal specification (see also (Merton, 1976)), double exponential (see also (Kou, 2002)), uniform model and then some less common implementations such as double Rayleigh and double uniform.

To illustrate how the CPS model can be used to increase the default correlation, first accept the definition of the 'default correlation over time' $\rho_{i,j}(t)$ stated in (III.2.14). Our objective is to analyze whether the CPS setup allows for stronger default correlations than the classical Itô, given that both setups populate approximately close term-structures of default probabilities. Assume that there are two names 1 and 2 whose term structure of default probabilities $t \in [0, T^*] \mapsto \mathbb{P}[\tau_i^* \leq t], i = 1, 2$ are observable. If we match both probability curves, can this match be achieved with a different default correlation $\rho_{1,2}(t)$? We have pre-generated two default-probability curves originating from Itô's processes. These have the solid lines in the Figure 19(a). Then, we have specified a CPS system with $\lambda = 2.0$ such that these probability curves are approximately matched by the CPS model. The 'matching' default probability curves under CPS have dashed lines in the Figure 19(a). This should intuitively suggest that the default correlation could possibly be the same in both the Itô and CPS specification of the intensities. We, however, found that there is a significant difference between the $\rho_{1,2}(t)$ under both setups. The CPS setup exhibits a much higher default correlation than that of Itô despite it produces nearly similar individual probabilities. This difference in correlations is well visible on the Figure 19(b). In both approaches we used a perfect correlation of the Wiener processes $\rho = 1$ so technically with the given set of parameters the default correlation is in the limit for the Itô setup.

Notice that in order to evaluate default correlation $\rho_{1,2}(t)$ using the expression (III.2.14) we need to compute not only (i) the individual probabilities $\mathbb{P}[\tau_i^* \leq t], i = 1, 2$ but also (ii) 'probability of the first default' $\mathbb{P}[\min[\tau_1^*, \tau_2^*] \leq t]$. Since all these probabilities can be written in terms of expectations $\mathbb{E}[\cdot]$ of the appropriate default indicators, then all the terms $\mathbb{P}[\tau_i^* \leq t], i = 1, 2$ and $\mathbb{P}[\min[\tau_1^*, \tau_2^*] \leq t]$ (or their survival complements) also satisfy a particular PDE (Itô) or PIDEs (CPS) of the class (III.2.18) which we both numerically evaluated. A numerical solution to the PIDE for the 'probability of the first default' $\mathbb{P}[\min[\tau_1^*, \tau_2^*] \leq 5]$ is shown in the Figure 20.

We summarize this section with the following findings. (i) Although a default in an intensity model is a jump process, the intensity itself can follow jumps without violating the general valuation principles. In turn a model with 'jump (default) induced by a jump-diffusion process (intensity)' still allows for a convenient numerical treatment. (ii) The valuation formula turns out to be a standard PIDE. (iii) The Common Poisson Shock specification allows to achieve a stronger default correlation than what is possible with Itô intensities. (iv) In CPS the change of the default correlation requires to adjust the

⁹⁷the common estimation problem is to identify what is a diffusion and what is a jump when a time series is observed. In this context it sometimes makes more sense to use such a specification which considers only a sudden movement of a particular magnitude as jumps otherwise the movement is a diffusion.

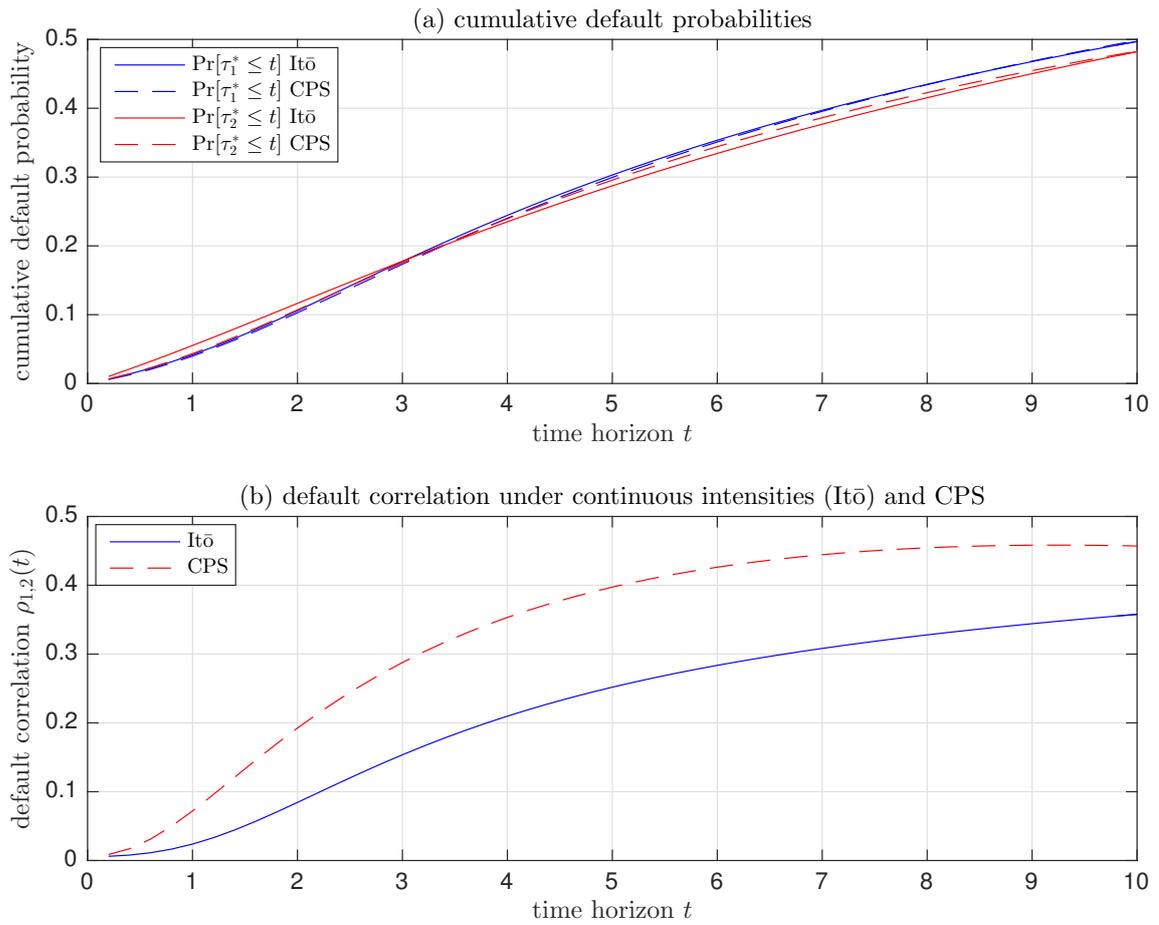


Figure 19: (a) individual cumulative default probabilities under the Itô and the CPS setup. (b) achieved default correlations under the Itô and the CPS setup.

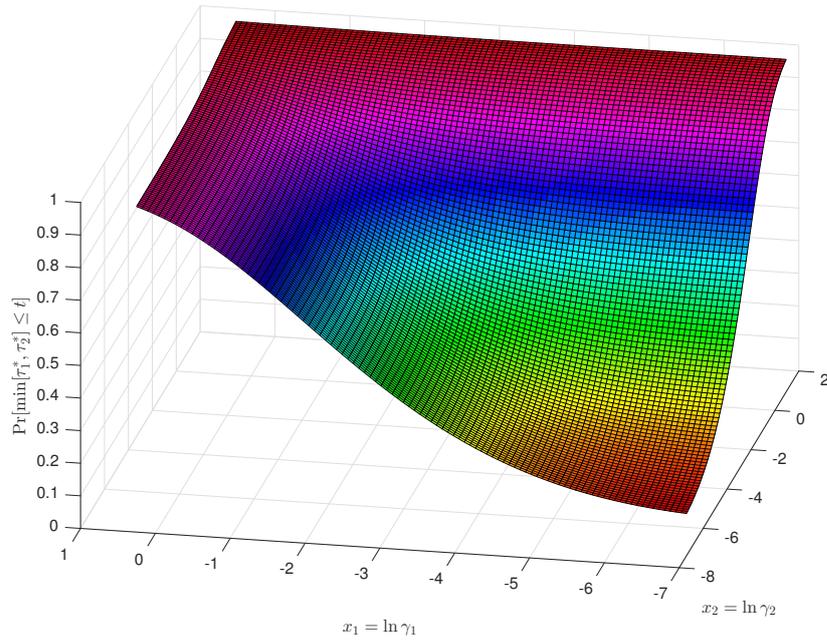


Figure 20: Numerical solution to PIDE for $\mathbb{E}[H_{i,j}(5)] = \mathbb{P}[\min[\tau_1^*, \tau_2^*] \leq 5]$ in the CPS model.

distributional properties of the intensities while in the Itô's models it does not.⁹⁸ (v) The CPS has only a one-dimensional convolution integral in the PIDE.

⁹⁸in Itô we can use the correlation of the Wiener processes ρ freely without any effect on the individual probabilities of default while in the CPS changing the jump components affects the probabilities of default.

III.3 A defaultable binomial tree

In the preceding long Section III.2 we have worked under the most general setup allowing for the stochasticity of all the factors involved and under this setup we have constructed customized pricing PDEs which we applied to various practical problems. Although the methodology was almost completely general, it still required relatively powerful numerical methods to obtain the correct results. Following our previous research paper (Kolman, 2015a) the aim of this section is to introduce a novel approach of pricing defaultable bonds which is based on a binomial tree approximation. The main essence of this approach is a *two-level tree* that supports absorbing default and non-default states which is perfectly suitable for pricing of defaultable securities, in particular bonds. The benefit of this structure is that the dimensionality of the tree structure is preserved, i.e. the number of the nodes in the tree structure grows with increasing time steps in the same way as it grows in a classical tree structure with no default feature. This makes the proposed two-level grid particularly attractive.

The basic mechanics of the default risk we adopt is that of (Jarrow and Turnbull, 1995) (henceforth shorthanded as 'JT') which indeed offers a flexible and intuitive framework for modeling the default risk. The main and critical property of Jarrow and Turnbull's framework is that they assume the Fractional Recovery of the Treasury Value (FRTV)⁹⁹ which makes the tree complete because regardless of the default or no default the payoff always takes place at the security's maturity T . For the modeling of the risk-free term-structure we have selected the Black-Derman-Toy (BDT) (Black et al., 1990) model which is known to be a well established methodology for modeling interest rates such that the initial yield curve is perfectly matched. Although the assumption of JT being in place is critical, the framework can support any kind of the interest-rate model (not necessarily just the BDT) which allows for tree approximation of interest rate evolution. Particularly attractive are the models based on the Heath-Jarrow-Morton framework (Heath et al., 1990, 1991, 1992) in the multi-factor setup, yet one must bear in mind that the HJM multi-factor models often populate non-recombining 'bushy trees' where the number of points at the time $t = i\Delta_t$ is equal to $(n_f + 1)^i$ where n_f denotes the number of factors in consideration and so the complexity of the tree grows quickly with increasing time step i . We recommend reading (Deventer, 2013, Section 6) on general construction of these HJM bushy trees and the reader can also visit the Appendix A.2. As the main objective is to explain the modeling environment rather than to investigate some particular issues, for the exposition here we make several simplifying assumptions which can later be relaxed.

III.3.1 The model setup

We assume that the model setup follows the ideas behind the Jarrow-Turnbull methodology and that the reduced-from theory outlined in the Section III.1 is in effect. The most critical assumptions in the JT methodology are that the intensity γ is a deterministic (or constant) function of time, default is an absorbing state and FRTV recovery policy is in place. We assume that the continuous world can be approximated by N equidistant time intervals

⁹⁹see the beginning of the Section III.2 for the overview of the common recovery policies.

of the length $\Delta_t = T/N$ such that $t_i = i\Delta_t, i = 0, \dots, N$ which means $t_0 = 0$ and $t_N = T$ and there are $(N + 1)$ time points in total.

Although it is not necessary, to make the exposition simple we assume that the term-structure of the one-period forward default probabilities is flat, i.e. we assume

$$\mathbb{Q}[t_i < \tau^* \leq t_{i+1} | \tau^* > t_i] = \pi \text{ for all } i.$$

In this view, the default intensity $\gamma(t)$ is a constant $\gamma(t) \equiv \gamma$ equal to

$$\gamma = -\frac{1}{\Delta_t} \ln(1 - \pi).$$

We can also conclude that under this discrete setup, the cumulative probability of default until time $t = t_n = n\Delta_t$ expressed in terms of the forward default probabilities is simply

$$\mathbb{Q}[\tau^* \leq t_n | \tau^* > 0] = 1 - \prod_{i=0}^{n-1} (1 - \mathbb{Q}[t_i < \tau^* \leq t_{i+1} | \tau^* > t_i]),$$

assuming the credit is not in default at the time $t = 0$. Since the forward default probability is assumed to be π for all the periods this expression reduces to

$$\mathbb{Q}[\tau^* \leq t_n | \tau^* > 0] = 1 - (1 - \pi)^n,$$

which makes the it very easy to be understood.

A critical assumption inherent to the JT framework is that of the FRTV recovery policy being in effect. According to this recovery specification, if the credit defaults the bondholders receive a recovery δ at the credit's maturity T *regardless of the time of default* τ^* . Since given default $\tau^* \leq T$ we always receive δ at T , this is conditionally on the default a *sure* amount and so the value of the defaulted bond at time τ^* is the risk-free discounted value of the recovery δ . In other words, denoting P the risky bond, we have

$$P(t, T) = \delta B(t, T), \tau^* \leq t \leq T.$$

Let us focus on the terminal payoff at time T . There the payoff can have only two states

$$P(T, T) = \begin{cases} \delta & \text{if } \tau^* \leq T \\ 1 & \text{if } \tau^* > T \end{cases}.$$

The default can thus be related to a hypothetical FX rate X in which the bond is denominated. The FX rate is

$$X(t) = \begin{cases} \delta & \text{if } \tau^* \leq t \\ 1 & \text{if } \tau^* > t \end{cases}, \quad (\text{III.3.1})$$

or written compactly as $X(t) = \mathbf{1}_{\{\tau^* > t\}} + \delta \mathbf{1}_{\{\tau^* \leq t\}}$. In terms of X the bond price P can

be expressed as

$$\begin{aligned}
P(t, T) &= B(t, T)\mathbb{E}^{\mathbb{Q}} [X(T) | \mathcal{H}(t)] \\
&= B(t, T)\mathbb{E}^{\mathbb{Q}} [\mathbf{1}_{\{\tau^* > T\}} + \delta\mathbf{1}_{\{\tau^* \leq T\}} | \mathcal{H}(t)] \\
&= \mathbf{1}_{\{\tau^* > t\}}B(t, T)\mathbb{E}^{\mathbb{Q}} [\mathbf{1}_{\{\tau^* > T\}} + \delta\mathbf{1}_{\{\tau^* \leq T\}} | \tau^* > t] + \mathbf{1}_{\{\tau^* \leq t\}}\delta B(t, T) \\
&= \mathbf{1}_{\{\tau^* > t\}}B(t, T)\mathbb{E}^{\mathbb{Q}} [1 - \mathbf{1}_{\{\tau^* \leq T\}} + \delta\mathbf{1}_{\{\tau^* \leq T\}} | \tau^* > t] + \mathbf{1}_{\{\tau^* \leq t\}}\delta B(t, T) \\
&= \mathbf{1}_{\{\tau^* > t\}}B(t, T)\mathbb{E}^{\mathbb{Q}} [1 + \mathbf{1}_{\{\tau^* \leq T\}}(\delta - 1) | \tau^* > t] + \mathbf{1}_{\{\tau^* \leq t\}}\delta B(t, T) \\
&= \mathbf{1}_{\{\tau^* > t\}}B(t, T) (\mathbb{Q}[t < \tau^* \leq T | \tau^* > t] (\delta - 1) + 1) + \mathbf{1}_{\{\tau^* \leq t\}}\delta B(t, T).
\end{aligned} \tag{III.3.2}$$

This formula is general, that is it supports both the defaulted and non-defaulted cases as seen from the time t . If we restrict the formula for P to only the non-defaulted cases, then the formula (III.3.2) reduces to

$$P(t, T) = B(t, T)\mathbb{E}^{\mathbb{Q}} [X(T) | \tau^* > t] = B(t, T) (\mathbb{Q}[t < \tau^* \leq T | \tau^* > t] (\delta - 1) + 1), \tag{III.3.3}$$

that is the present value of the defaultable bond $P(t, T)$ prior to default is equal to the risk-free discounted risk-neutral probability of receiving either just the recovery δ or the full bond notional 1 at the bond's maturity T . The formula (III.3.3) is a key component of the modeling framework as it connects the evolution of the risk-free rate and the probability of default with the pre-default value of a defaultable bond.

Having discussed the credit risk process it remains to specify the risk-free rate dynamics. As outlined above, we have selected the Black-Derman-Toy model which is completely sufficient for the illustration of the model as it is designed to be used for binomial tree approximation of evolution of the risk-free rate. The generalized \mathbb{Q} -dynamics of the short-rate $r(t)$ in BDT is given as

$$d \ln r(t) = \left(\theta(t) + \frac{\sigma'(t)}{\sigma(t)} \ln r(t) \right) dt + \sigma(t) dW^{\mathbb{Q}}(t).$$

For the exposition here, it is sufficient¹⁰⁰ to use the constant-volatility term $\sigma(t) \equiv \sigma$ which implies a simplified equation for the dynamics of $\ln r$

$$d \ln r(t) = \theta(t) dt + \sigma dW^{\mathbb{Q}}(t).$$

The deterministically dependent term $\theta(t)$ serves to match the observed yield curve. It is also easy to see that this specification of the interest-rate is actually a Ho-Lee dynamics (Ho and Lee, 1986) where $dr(t)$ is replaced by $d \ln r(t)$. For the discrete binomial representation, let

$$r_{i,j} = r_j(i\Delta_t) \equiv r_j(i\Delta_t, (i+1)\Delta_t),$$

denote the short rate observed at $t = t_i = i\Delta_t$ in position j ,¹⁰¹ paying out the interest one period later at $t = t_{i+1} = (i+1)\Delta_t$. To construct the tree, it is given that the interest rate

¹⁰⁰also, when $\sigma(t)$ is left to be deterministic function of time it might cause the short rate to explode.

¹⁰¹we can refer to j as to the 'level' of the risk free rate. In our setup j can be positive, zero or a negative integer.

must go up or down with probability $p = 0.5$ in every step and that the tree is recombining. The index j of level of the interest rate can actually be seen as a function of the time index i such that the range of j at the time step $t = i\Delta_t$ is $j(i) = -i, -i + 2, \dots, i - 2, i$. An example of a three-period BDT model is shown in Figure 21. Consider a sure unit cash-

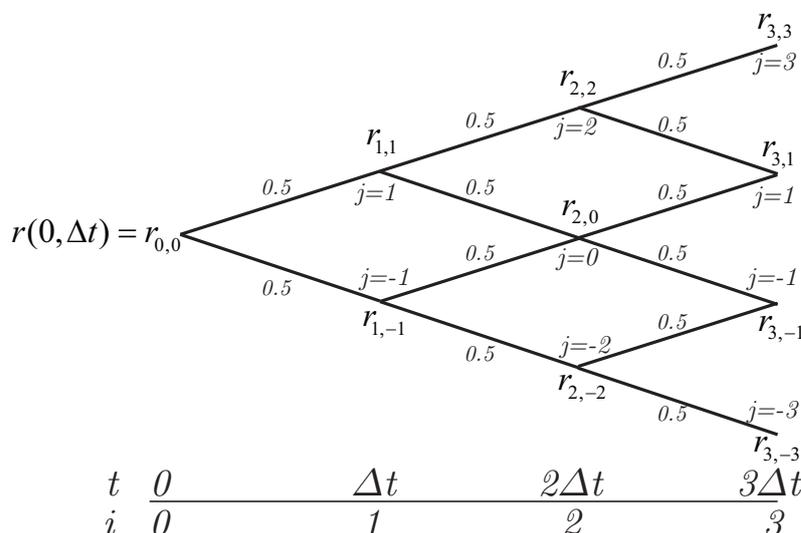


Figure 21: Evolution of risk-free rate r in a recombining three-period binomial interest-rate tree.

flow to be received at time $(i + 1)\Delta_t$ and assume that the risk-free rate is in position j at time $i\Delta_t$. Then, as seen from time $i\Delta_t$ (and position j) the value of this unit cash flow is

$$d_{i,j} = \frac{1}{1 + r_{i,j}\Delta_t}. \quad (\text{III.3.4})$$

In other words, $d_{i,j}$ is the $(i\Delta_t)$ -value of sure unit cash flow received one period later given the interest rate is in position j at the time $(i\Delta_t)$. This simple expression for $d_{i,j}$ can easily be inverted so that knowing the one-period discount $d_{i,j}$ uniquely determines the one-period rate $r_{i,j}$. In BDT the short-rate r at time $i\Delta_t$ has the following pre-specified functional form with respect to the level j

$$r_{i,j} = \lambda(i) \exp\left(\sigma j \sqrt{\Delta_t}\right), \quad (\text{III.3.5})$$

where $\lambda(i)$ is a time-dependent parameter¹⁰² which is to be found iteratively. The last essential component of the BDT model is the Arrow-Debreu security Q which has the meaning of being a *discounted expectation*. In other words, the present value ($t = 0$) of reaching the point (i, j) at which 1 is received, is equal to $Q_{i,j}$.

If we fix i and take a look at $Q_{i,j}$ then it is clear that they must sum to $B(0, i\Delta_t)$ because Q s in terms of probability fully describe all states at time $i\Delta_t$ and these probabilities are discounted to $t = 0$. Put another way exactly one j state at time $t = i\Delta_t$ of all possible states *has to* occur and so the probabilities covered by Q must be complete and sum to 1.

¹⁰²Jamshidian discovered that $\lambda(i)$ is a median of the short rate at $t = i\Delta_t$.

For arbitrary i we write

$$\sum_{j(i)} Q_{i,j} = B(0, i\Delta_t), \quad (\text{III.3.6})$$

where we explicitly wrote $j(i)$ as a state j given the time point i . Consider a zero-bond maturing at $(i+1)\Delta_t$. By definition of the discount factor in (III.3.4) we can discount¹⁰³ its sure cash-flow of 1 to time $i\Delta_t$ and then use (III.3.6). This implies that

$$B(0, (i+1)\Delta_t) = \sum_{j(i)} Q_{i,j} d_{i,j}.$$

Substituting for $d_{i,j}$ from (III.3.4) and subsequently $r_{i,j}$ from (III.3.4) we have

$$B(0, (i+1)\Delta_t) = \sum_{j(i)} \frac{Q_{i,j}}{1 + \lambda(i) \exp(\sigma j \sqrt{\Delta_t}) \Delta_t}. \quad (\text{III.3.7})$$

Thus knowing $Q_{i,j}$ and having the quoted price $B(0, (i+1)\Delta_t)$ allows to infer $r_{i,j}$. The last necessary component is the structure of $Q_{i,j}$. In BDT the $Q_{i,j}$ is defined in terms of forward induction

$$Q_{i,j} = \begin{cases} 0.5Q_{i-1,j-1}d_{i-1,j-1} & \text{if } j = i \\ 0.5Q_{i-1,j-1}d_{i-1,j-1} + 0.5Q_{i-1,j+1}d_{i-1,j+1} & \text{if } j \neq i, -i \\ 0.5Q_{i-1,j+1}d_{i-1,j+1}, & \text{if } j = -i \end{cases} \quad (\text{III.3.8})$$

Thus the construction of BDT resides in several relatively easy steps (see also (Clewlow and Strickland, 1998, Section 8.2)) :

1. Choose some finite horizon T and choose N as a number of time steps. This will yield a time step of the length Δ_t .
2. Select a candidate for the short rate $r(t) = r(i\Delta_t, (i+1)\Delta_t)$ and calculate its volatility σ from the historical data or by matching the market prices of interest-rate sensitive securities.
3. Start with $i = 0$. $Q_{0,0} = 1$ because the value of Arrow-Debreu security in the current state is 1. $r_{0,0} = r(0, \Delta_t)$ is observable which gives the discount factor $d_{0,0} = \frac{1}{1+r_{0,0}\Delta_t}$.
4. Set $i := i + 1$ and apply the forward induction (III.3.8) to obtain $Q_{i,j}$ for all j given i .
5. Use bond price $B(0, (i+1)\Delta_t)$ to iteratively¹⁰⁴ obtain $\lambda(i)$ via (III.3.7).
6. Having $\lambda(i)$ we can calculate $r_{i,j}$ by (III.3.5) and $d_{i,j}$ using (III.3.4) and save them to the tree.
7. Repeat steps 3 to 6 until $i \leq N$.

¹⁰³notice that we do not need to analyze what will happen in the upcoming period if we are in (i, j) because we are always guaranteed that in $(i+1)\Delta_t$ we receive 1 no matter what is the value of $r_{i+1,j}$.

¹⁰⁴Newton-Raphson with a reasonable positive initial value such as 0.02 is recommended to be used.

8. Using the tree of discount factors $d_{i,j}$ create a tree of risk-free bond prices as $B_{i,j} = d_{i,j}(0.5B_{i+1,j-1} + 0.5B_{i+1,j+1})$ and go backward (thus with decreasing i) unless $i = 0$.

As it is apparent from the BDT construction algorithm, the whole procedure actually requires to create four trees of which some are auxiliary. These four trees are the tree of evolution of r , tree for discount factors d , tree of Arrow-Debreu prices Q and finally a tree of bond prices B .

Above, we have outlined some specifics of the default/survival processes in our model and we have also described the construction of the BDT binomial tree for the short rate/bond values. We are thus equipped to model a defaultable bond as a combination of risk-free rate (or bond) evolution process and default/survival process. The methodology we propose is based on a very simple principle of separation of the defaulted and nondefaulted bond prices. Therefore, all possible states of a defaultable claim can be captured by a two-level tree.

Consider a three-period default-free bond B whose price process is fully determined by the evolution of short rates in BDT. The price process of such a bond is captured by the Figure 22. Secondly, consider a three-period process of the virtual FX rate X which has already been discussed earlier in this text and formally defined in (III.3.1). The conditional expectation of X is

$$\mathbb{E}^{\mathbb{Q}}[X(T)|\mathcal{G}(t)] = \mathbb{E}^{\mathbb{Q}}[X(T)|\mathcal{H}(t)] = \delta \mathbf{1}_{\{\tau^* \leq t\}} + \mathbf{1}_{\{\tau^* > t\}} \mathbb{E}^{\mathbb{Q}}[X(T)|X(t) = 1],$$

where T is the 'terminal time', in this case the payoff time of a bond. Process X and its conditional \mathbb{Q} -expectation $\mathbb{E}^{\mathbb{Q}}[X(3)|X(t)]$ is visually represented in the Figure 23.

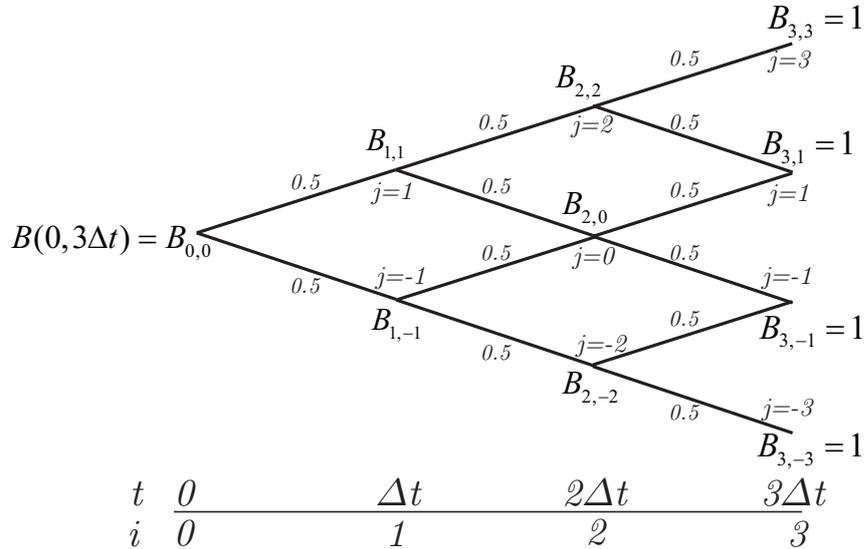


Figure 22: Evolution default-free bond in a three-period interest-rate tree. The evolution is consistent with the risk-free rate evolution in the Figure 21.

We briefly discuss the two processes given in the Figure 23. The upper scheme in this figure is a process of X as observed at the time points $0, \Delta t, 2\Delta t, 3\Delta t$. At the outset (time

$t = 0$) the virtual FX rate $X(0) = 1$, i.e. the bond is *not* in default. Looking one step ahead (time $t = \Delta_t$) the bond might have defaulted within $(0, \Delta_t]$ with a probability π , thus $X(\Delta_t) = \delta$ or has not defaulted with the complementary probability $(1 - \pi)$ and so $X(\Delta_t) = 1$. If it has not defaulted then it stays on the upper (black) line of survival. If it has defaulted it falls down to the red-line default area. Knowing the state of $X(\Delta_t)$ gives us possible outcomes of $X(2\Delta_t)$. If the bond survived till Δ_t , then again it might default with probability π in the time interval $(\Delta_t, 2\Delta_t]$ inducing $X(2\Delta_t) = \delta$. If it survived till Δ_t and also survived till $2\Delta_t$ then clearly $X(2\Delta_t) = 1$. Finally, if it were in default at Δ_t (i.e. $X(\Delta_t) = \delta$) then it *must* stay in default in the upcoming time point $2\Delta_t$ because the default is an absorbing state (i.e. $X(2\Delta_t) = \delta$). Therefore, in the red-line area the probability of staying in default is 1.

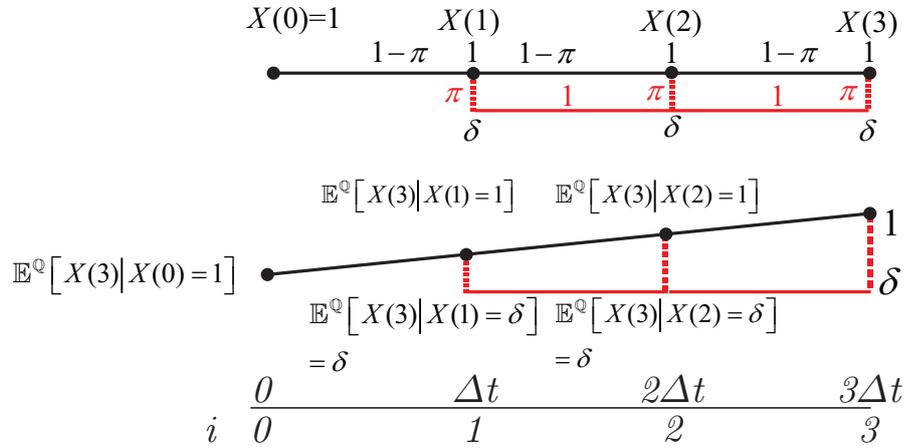


Figure 23: Visualization of (i) process of immediate 'FX rate', (ii) process of expected payoff in a three-period model with constant default probability π .

The lower scheme shows a process of conditional expected value of the virtual FX rate X . We start at $t = 0$ and we know that the bond has to survive till the maturity $3\Delta_t$ to pay 1 (otherwise it pays just $\delta < 1$ at $t = 3\Delta_t$). Therefore as of the very beginning it is relatively likely that the bond will not survive until $3\Delta_t$ because there are 3 possibilities to jump to default in the future - in the intervals $(0, \Delta_t]$, $(\Delta_t, 2\Delta_t]$ or $(2\Delta_t, 3\Delta_t]$. To sum up at the time $t = 0$ the expected value of $X(3\Delta_t)$ is quite low. If the bond defaults during the first period $(0, \Delta_t]$ then it appears down in the red-line area and the expected value of $X(3\Delta_t)$ at time $t = \Delta_t$ is δ because, again, default is an absorbing state. Otherwise it has survived through the first period $(0, \Delta_t]$ and the Δ_t -expected value must take into account that it is more likely that $X(3\Delta_t)$ will be 1 than it was at $t = 0$. We could move ahead to the future with the same logic but these future steps are obvious. It is clear from the principle that the survival line is increasing in time while the default line is fixed at δ .

If we combine the default-free term structure shown in the Figure 22 with the bottom scheme of Figure 23 we have a two-level arbitrage-free tree that captures the dynamics of the defaultable bond when default and interest-rates affect the price of the bond. The

absence of arbitrage is ensured through the existence of the martingale probabilities¹⁰⁵ \mathbb{Q} for both the evolution of the risk-free rate and the evolution of default/survival states. This is perfectly consistent with (III.3.3) where the defaultable bond price P is just simply a product of the default-free bond price B and the expectation of the bond's payoff X when not in default and as a product of the recovery δ and the risk-free bond B if in default. We therefore have a defaultable bond price tree P as shown in the Figure 24.

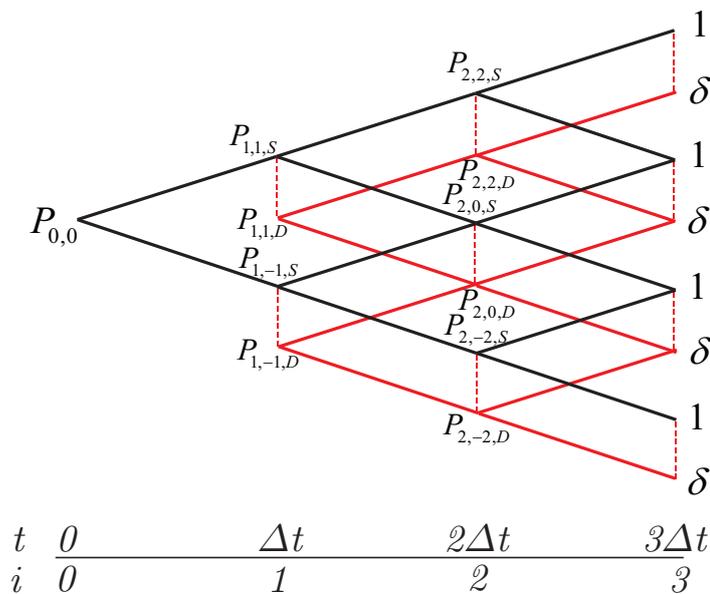


Figure 24: Tree of defaultable bond prices $P_{i,j,z}$, $z \in \{S, D\}$, where S means Survival and D means Default.

The initial point $P_{0,0}$ is the bond price quote that we currently observe. Looking at $t = \Delta t$ four scenarios are possible. If the bond survived in $(0, \Delta t]$, it stays in the black-line area where all the states are denoted $P_{i,j,S}$. Otherwise, in the case of default, it will jump to the red-line area where states are denoted $P_{i,j,D}$. In any case, and independently on the default/survival the defaultable bond price P will be driven by the state of the risk-free rate which might be in the first time-step $i = 1$ either $r_{1,1}$ or $r_{1,-1}$.

By the product scheme (or from (III.3.2)) we can easily see that the price P of $N\Delta t$ -maturity defaultable bond is given by

$$P_{i,j,z} = \begin{cases} B_{i,j} \left((1 - \pi)^{N-i} + \delta \left[1 - (1 - \pi)^{N-i} \right] \right) & \text{if } z = S \\ B_{i,j} \delta & \text{if } z = D \end{cases}. \quad (\text{III.3.9})$$

The two-level P bond price tree is an easy and tractable tool for modeling derivatives on defaultable bonds. The later chapters cover this topic in a greater detail.

III.3.2 Approximating extension to coupon-bearing bonds

So far we have been assuming that the underlying bond P is a zero-coupon bond. This is also the key assumption of the Jarrow-Turnbull model, which we consider as a benchmark.

¹⁰⁵See the First Fundamental Theorem (Theorem I.1.10).

We, however, show that the structure we propose *allows* for the coupons to be implemented which greatly increases the applicability because the majority of bonds bear coupons. The idea is simple and intuitive. A coupon payment paid at T_m (not necessarily paid exactly at any of the time points which arise from discretizing the time axis to $t_i = i\Delta_t$) is modeled as a 'small' zero-bond paying out the coupon if not in default at the payment date and paying nothing when in default. This is in accordance with the general principle of the recovery payoff which is related to the bond's notional only and not to the promised coupon payments. Then we find a bucket U^k to which the cash coupon will be added. The bucket U^k (with a hypothetical maturity t_k) collects the coupon cash-flow streaming in the interval $[t_k, t_{k+1})$. The logic behind this cash-flow bucketing is that as of t_k the bond still trades with the coupon paid at $T_m \in [t_k, t_{k+1})$ but at t_{k+1} the bond already trades without the coupon, therefore it is more natural to map the coupon into the time point t_k where the coupon is still part of the bond. Having bucketed all the coupons, we treat every bucket as a risky-zero bond that pays U^k at time t_k when not in default and zero otherwise. We construct the same tree for every bucket U^k (each maturing at different t_k) as for the P above and once this is done, we can add all U^k to the survival level tree for P . We describe the whole process algorithm-wise (we assume the bond matures at time s):

1. For every time index $k = 1, \dots, N_s$ step created for P create a coupon cash-flow bucket U^k collecting the coupon cash-flow streaming in the interval $[t_k, t_{k+1})$. (U^{N_s} only holds cash-flow streaming at $t = s$).
2. To every cash-flow bucket U^k add all the cash-coupons which are paid in $[t_k, t_{k+1})$.
3. Regard U^k as a zero-bond P maturing at $t_k = k\Delta_t$, having no recovery ($\delta = 0$).
4. Construct for every bucket U^k tree $U_{i,j}^k$ by using (III.3.9) where $P_{i,j,S} = U_{i,j}^k$, $z = S$, $\delta = 0$. This means

$$U_{i,j}^k = B_{i,j}^k (1 - \pi)^{k-i}, 0 \leq i \leq k,$$

where $B_{i,j}^k$ is a BDT tree created for a default-free zero bond maturing at $t_k = k\Delta_t$.

5. Add the trees $U_{i,j}^k$ to $P_{i,j,S}$ by overriding the previous values of $P_{i,j,S}$, that is in the sense of the 'algorithmic notation'

$$P_{i,j,S} := P_{i,j,S} + \sum_{k=1}^{N_s} U_{i,j}^k. \quad (\text{III.3.10})$$

This algorithm above produces the 'dirty' price of the coupon bond P . Further adjustments would have to be made to model P as a clean price.

III.3.3 European options on defaultable bonds

Pricing of the European options on the defaultable bonds using the defaultable tree is done in a natural way and essentially does not differ from the standard binomial option tree method. Even though two levels actually enter the pricing here, it is

not hard to reflect these levels in the option price because the option price is only determined by the 'terminal' states on the tree and not on earlier nodes, hence the option has no path-dependency. In what follows assume $s \geq T$ is the maturity of the underlying bond and T is the maturity of the option in question which we assume is a call option.

At the option's maturity T , we observe the payoff

$$V_{N_T,j,z} = \begin{cases} \max[P_{N_T,j,S} - K, 0] & \text{if } z = S \\ \max[P_{N_T,j,D} - K, 0] & \text{if } z = D \end{cases}.$$

Since the survival ($z = S$) and the default ($z = D$) states occur with the probabilities $(1 - \pi)^{N_T}$ and $1 - (1 - \pi)^{N_T}$, respectively, we have the the default/survival probability-weighted payoff

$$V_{N_T,j} = (1 - \pi)^{N_T} \max[P_{N_T,j,S} - K, 0] + \left(1 - (1 - \pi)^{N_T}\right) \max[P_{N_T,j,D} - K, 0].$$

Substitution for P from (III.3.9) yields the T -payoff

$$\begin{aligned} V_{N_T,j} &= (1 - \pi)^{N_T} \max[B_{N_T,j} \left((1 - \pi)^{N_s - N_T} + \delta \left[1 - (1 - \pi)^{N_s - N_T} \right] \right) - K, 0] \\ &+ \left(1 - (1 - \pi)^{N_T} \right) \max[B_{N_T,j} \delta - K, 0]. \end{aligned} \quad (\text{III.3.11})$$

Having the payoff expression at the option's maturity $N_T \Delta_t = T$, we obtain the current price of the option at the point $(0, 0)$ by the usual backward iteration in time. Starting from $i = N_T - 1$, we compute

$$V_{i,j} = \frac{1}{2} d_{i,j} (V_{i+1,j+1} + V_{i+1,j-1}), i \geq 0. \quad (\text{III.3.12})$$

The approach above already reflects the probabilities $(1 - \pi)^{N_T}$, $1 - (1 - \pi)^{N_T}$ of achieving the terminal nodes on the survival/default level. It has the benefit that no other tree than just for V needs to be created.¹⁰⁶ An alternative but numerically equivalent approach is to reflect these probabilities 'as-you-go' when iterating backwards through the tree and we present this here for the sake of completeness. This approach will then be more natural for pricing of the American options. We firstly create a tree V^δ which models the value of the option on the *defaulted* level. The tree V^δ is established as follows:

$$V_{i,j}^\delta = \begin{cases} \max[B_{N_T,j} \delta - K, 0] & \text{if } i = N_T \\ \frac{1}{2} d_{i,j} (V_{i+1,j+1}^\delta + V_{i+1,j-1}^\delta) & \text{if } 1 \leq i < N_T \end{cases}. \quad (\text{III.3.13})$$

Notice here that $V_{0,0}^\delta$ is not computed because it will never be used for the fact that at time $t = 0$ the underlying bond is not in default. After this tree of the option prices on the defaulted level is created, we can construct the option price tree V as

$$V_{i,j} = \begin{cases} \max[B_{N_T,j} \left((1 - \pi)^{N_s - N_T} + \delta \left[1 - (1 - \pi)^{N_s - N_T} \right] \right) - K, 0] & \text{if } i = N_T \\ \frac{1}{2} d_{i,j} \left[\pi (V_{i+1,j+1}^\delta + V_{i+1,j-1}^\delta) + (1 - \pi) (V_{i+1,j+1} + V_{i+1,j-1}) \right] & \text{if } 0 \leq i < N_T \end{cases}. \quad (\text{III.3.14})$$

¹⁰⁶clearly, we need the BDT trees.

The main difference between (III.3.11)–(III.3.12) and (III.3.13)–(III.3.14) is that in the former case, we built in the default-survival probabilities directly to the terminal nodes and applied a combined default/survival option payoff. This simplifies the scheme such that only one tree V is required. The latter case is more complicated for the construction (we have to create the extra tree V^δ) but represents the binomial structure in a more natural way and more importantly allows for an easy extension to the American options.

III.3.4 American options

The previous section presented an option pricing scheme for pricing the European options on defaultable bonds. This can be extended to American options in a standard way. In this case, both the time of default and the level of the interest rates *do* influence the optimal exercise and thus capturing both these factors in the scheme is the point of interest. The immediate exercise of the option can conveniently be built into the option trees (III.3.13)–(III.3.14) by adding the early exercise feature. Let us denote the immediate exercise payoff

$$\varphi_{i,j,z} = \max [P_{i,j,z} - K, 0], \text{ where as usual } z \in \{S, D\}.$$

The tree V^δ reflecting the price of the option on the defaulted level is adjusted to

$$V_{i,j}^\delta = \begin{cases} \max [B_{N_T,j} \delta - K, 0] & \text{if } i = N_T \\ \max \left[\frac{1}{2} d_{i,j} (V_{i+1,j+1}^\delta + V_{i+1,j-1}^\delta), \varphi_{i,j,D} \right] & \text{if } 1 \leq i < N_T \end{cases}. \quad (\text{III.3.15})$$

Once this tree is set up, we apply the adjusted version of (III.3.14) which is

$$V_{i,j} = \begin{cases} \max [B_{N_T,j} \left((1 - \pi)^{N_s - N_T} + \delta \left[1 - (1 - \pi)^{N_s - N_T} \right] \right) - K, 0] & \text{if } i = N_T \\ \left(\frac{d_{i,j}}{2} \left[\pi (V_{i+1,j+1}^\delta + V_{i+1,j-1}^\delta) + (1 - \pi) (V_{i+1,j+1} + V_{i+1,j-1}) \right], \varphi_{i,j,S} \right)^+ & \text{if } i < N_T \end{cases}, \quad (\text{III.3.16})$$

where it was necessary to switch the notation of $\max[\cdot]$ to $(\cdot)^+$ to fit everything into one line. In (III.3.16), for $i < N_T$ we applied the logic that the bond is in a survival state at time $t = i\Delta_t$ where the immediate exercise $\varphi_{i,j,S}$ can be applied or the investor can hold the option for one more period such that the underlying bond can in the future time step either jump to default with probability π or survive with the probability $1 - \pi$.

In what follows we are about to find a pricing methodology for the barrier options on defaultable bonds. To make the exposition clearer, we firstly recall some important principles applied in the pricing of barrier stock options.

III.3.5 Revisiting barrier options on a stock. The binomial tree approach

In this supplemental section we show a general methodology for pricing of stock barrier options. The motivation behind this is that the two-level defaultable tree model uses a similar technique to that of for the stock barrier options but in a different framework. The proposed methodology is suitable because it makes no assumptions about the motion of the underlying if the underlying is captured by one level only. Thus, for example, a barrier option tree for an option on a default-free bond would require us to take exactly

the same steps as if we valued a barrier option on a stock S .

Denote $S_{i,j}$ the stock price in a given (i, j) node. We firstly consider the case of **up-and-in** option. In such a case the option is activated if and only if the underlying stock S hits some 'activation' barrier h . In turn, the up-and-in stock option requires us to create *two* option trees. One tree V^A which assumes that the option has already been activated in some virtual prior node, and another tree V which contains values of the option taking into account activation or no activation. On V we start with $i = N_T$ as follows

$$V_{N_T,j} = \begin{cases} \varphi_{N_T,j} & \text{if } S_{N_T,j} \geq h \\ 0 & \text{if } S_{N_T,j} < h \end{cases}. \quad (\text{III.3.17})$$

Then iterate backwards from $i = N_T - 1$ to $i = 0$ by taking the steps

$$V_{i,j} = \begin{cases} V_{i,j}^A & \text{if } S_{i,j} \geq h \\ d_{i,j}(pV_{i+1,j+1} + (1-p)V_{i+1,j-1}) & \text{if } S_{i,j} < h \end{cases}, \quad (\text{III.3.18})$$

where p is a \mathbb{Q} -probability of S going up in the following period.

To make the exposition complete, we add also a pricing scheme for the **up-and-out** barrier option. It requires no additional tree (unlike the 'in' option) and so all the steps are applied to a single pricing tree V only.

At the option's maturity T , that is at the time point $i = N_T$ we set

$$V_{N_T,j} = \begin{cases} 0 & \text{if } S_{N_T,j} \geq h \\ \varphi_{N_T,j} & \text{if } S_{N_T,j} < h \end{cases}. \quad (\text{III.3.19})$$

Logically, this is the exact opposite of up-and-in in (III.3.17). Again, going back from $i = N_T - 1$ to $i = 0$ we need to compute

$$V_{i,j} = \begin{cases} 0 & \text{if } S_{i,j} \geq h \\ d_{i,j}(pV_{i+1,j+1} + (1-p)V_{i+1,j-1}) & \text{if } S_{i,j} < h \end{cases}. \quad (\text{III.3.20})$$

For a better understanding of the problem we illustrate one sample case by a numerical example. We use CRR (Cox et al., 1979) stock tree with parameters $S(0) = S_{0,0} = 10, \sigma = 0.3, r = 0.05, K = 11, h = 11.8, T = 1, N_T = 3, \Delta_t = 0.3333$ and we price both up-and-in and up-and-out put options. The tree of prices of S is depicted in the Table 8 (left). The 'activated' tree¹⁰⁷ V^A is shown in the same table on the right.

¹⁰⁷this is actually an ordinary vanilla plain European option tree.

j/i	0	1	2	3
3				16.8138
2			14.1398	
1		11.8911		11.8911
0	10.0000		10.0000	
-1		8.4097		8.4097
-2			7.0722	
-3				5.9475

j/i	0	1	2	3
3				0.0000
2			0.0000	
1		0.6137		0.0000
0	1.4972		1.2608	
-1		2.4496		2.5903
-2			3.7460	
-3				5.0525

Table 8: CRR stock tree (left) and 'activated' tree V^A (right).

The Table 9 then holds the barrier-adjusted values of the barrier option V . One can see from the Table 9 that the up-and-in is much less valuable than the up-and-out. This is because the stock S must firstly move up (which gives prerequisites for landing out-of-the money) otherwise it would not activate. up-and-out is much more valuable because the knock out barrier level lies in the same direction as landing out-of-the-money anyway. It is important to note that vanilla plain European option can be replicated by a pair of barrier options of which one is 'in' and the other is 'out' (both either of the kind 'down' or 'up'), since the options in the pair are mutually complementary. It must therefore follow that

$$\text{Vanilla plain European} = \text{barrier option 'in'} + \text{barrier option 'out'}. \quad (\text{III.3.21})$$

In our example barrier up-and-in is exactly 0.304840823, barrier up-and-out is 1.192315261 (both numbers taken from the Table 9) and Vanilla plain European is 1.497156084 (from the Table 8 (right)) which shows that the parity (III.3.21) holds.

j/i	0	1	2	3
3				0.0000
2			0.0000	
1		0.6137		0.0000
0	0.3048		0.0000	
-1		0.0000		0.0000
-2			0.0000	
-3				0.0000

j/i	0	1	2	3
3				0.0000
2			0.0000	
1		0.0000		0.0000
0	1.1923		1.2608	
-1		2.4496		2.5903
-2			3.7460	
-3				5.0525

Table 9: up-and-in option tree (left) and up-and-out option tree (right).

III.3.6 Barrier options

Barrier options are much more complicated instruments compared to the European and the American options. The complication with the barrier options is obvious - there is a complete path dependency of the option price which may (or may not) be knocked-in/out throughout its life. As the underlying can attain values in two levels (default and non-default) this makes the pricing problem even more complex and it is obvious that a robust methodology is required to capture all the possible states that can occur.

The complexity of the valuation can be illustrated by considering the following example. Assume a put option up-and-in and the activation barrier h which is a number that can only be reached on the non-defaulted price level $P_{i,j,S}$. The option is likely to be (deeper) in-the-money on the defaulted level since there the bond will pay only the recovery δ . Yet, when we analyze the values on the defaulted surface we need to verify if the option was previously activated by hitting h or not. Therefore, what we need to incorporate in the model is an explicit path-dependency because it does matter when and where the underlying jumped to default, and also if the barrier h was hit or was not. Also notice that the underlying cannot jump back (and possibly activate the option) to the survival tree.

The key concept that we have developed for the analysis of the path dependence on the defaultable tree is something that we call the *Default Entry Point* (DEP). We denote DEP (i^*, j^*) as κ_{i^*, j^*} .¹⁰⁸ The pricing of the barrier options will essentially be based on the idea of a decomposition of the whole grid into all possible DEPs κ_{i^*, j^*} of which each has some value capturing the default feature of the bond. Sum of these values adjusted for the default probabilities and the effect of the barrier will eventually yield the default-price contributor to the overall barrier option value. We start with a Definition.

Definition III.3.1 (Default Entry Point (DEP))

Default entry point κ_{i^, j^*} is a point (i^*, j^*) for which the following holds*

$$\tau^* = i^* \Delta_t, \tag{III.3.22}$$

and where j^ has no influence on the default state and determines a specific grid position only.*

Note here that j^* has no effect on the default/survival because the hazard γ is a deterministic function of time and in turn also independent of r (this is the only market process in the system; see the Section III.1.1 where this is discussed in a greater detail). Clearly, by the structure of the binomial tree we can also say postulate the following simple Lemma.

Lemma III.3.1 (Number of DEPs)

Every binomial two-level tree has exactly $\frac{1}{2}N_s(N_s + 3)$ DEPs κ_{i^, j^*} .*

This trivial lemma ensures that default is possible everywhere on the grid except for the node $(0, 0)$ because the underlying bond starts from the survival state at the outset.

The technique of pricing of the barrier options which we propose is based on a simple idea. We have some survival states of the underlying bond (upper level of the binomial tree). The contribution of the survival surface is treated as being a single unit, i.e. we do *not* decompose the survival states into smaller units. Quantification of the price contribution of the default states (and the default probabilities) to the overall option value is, however, a more challenging task to be done. Therefore, to integrate the contribution of the default states into the total option price, we propose the decomposition of the whole grid into a set of all possible DEPs and determine the price contribution of these DEPs separately.

¹⁰⁸we use the asterisk mark to denote some fixed values of (i, j) .

Thus, the DEPs are treated as independent units. This, in other words means to construct a barrier binomial tree (or two trees in the case of knock-in option) for every DEP. This DEP has to be, however, 'lifted-up' to the survival surface which means also to reflect the probability of default at $t = i^* \Delta_t$ for a given DEP (i^*, j^*) .

Denote $\kappa_{i^*, j^*, i, j}$ the $t = i \Delta_t$ -value of the DEP κ_{i^*, j^*} as seen from the survival level. We define its (i^*, j^*) -value using the following Lemma.

Lemma III.3.2 (Initial survival-level value of DEP)

The $t = i^* \Delta_t$ -value of DEP κ_{i^*, j^*} as seen from the survival level is

$$\kappa_{i^*, j^*, i^*, j^*} = (1 - \pi)^{i^* - 1} \pi V_{i^*, j^*}^\delta,$$

where V_{i^*, j^*}^δ is stated in (III.3.13) (evaluated at $(i, j) = (i^*, j^*)$). DEP $\kappa_{i^*, j^*, i, j}$ becomes default-free $\forall i < i^*$.

Since for the pricing it is critical to determine the current (0,0)-value of the DEP, we need to introduce also the backward recursion for the prior time steps in the binomial tree.

Lemma III.3.3 (Survival-level value of DEP in previous time-points)

The value of κ_{i^*, j^*} in the nodes $i < i^*$ is

$$\kappa_{i^*, j^*, i, j} = \frac{1}{2} d_{i, j} (\kappa_{i^*, j^*, i+1, j+1} + \kappa_{i^*, j^*, i+1, j-1}). \quad (\text{III.3.23})$$

If $\kappa_{i^*, j^*, i+1, j+1}$ or $\kappa_{i^*, j^*, i+1, j-1}$ does not exist (is undefined) for a given κ_{i^*, j^*} then 0 is assumed for $\kappa_{i^*, j^*, i+1, j+1}$ or $\kappa_{i^*, j^*, i+1, j-1}$, respectively.

DEP κ_{i^*, j^*} is thus a just a mathematical construct which provides a price contribution to the overall option value by taking into account the jump into default exactly in (i^*, j^*) . The best way to look at κ_{i^*, j^*} is to assume that it is a 'quasi option'. The price contribution is then computed when the value of κ_{i^*, j^*} is rolled-back to the current time $t = 0$ from the initial future 'default-entry' position (i^*, j^*) .

It is also useful to introduce a superscript $k = 1, 2, \dots, \frac{1}{2} N_T (N_T + 3)$ to index the given DEP κ_{i^*, j^*} . For example κ^1 might mean the point $\kappa_{1, -1}$ so that a sum of all DEPs $\sum_k \kappa_{i^*, j^*, 0, 0}^k$ at the point (0,0) can be defined. At this stage we know how to get DEPs κ_{i^*, j^*} to the $t = 0$ -node (0,0). Yet, we still have not taken into account the barrier feature h . As pointed above, the κ_{i^*, j^*} can be treated as an (quasi) option. This point of view suggests using a similar technique for each DEP as for the stock barrier options revisited in the Section III.3.5 because in view of the Lemma III.3.2 the DEP is default-free and so we model this quasi option on the survival surface only. To ease the complexity, let us make the following, mild assumption.

Assumption III.3.1 (Separation of default states by h)

The values of the underlying P and and the barrier h are such that

$$P_{i, j, D} < h,$$

for all i, j .

This assumption effectively ensures that the effect of the barrier affects only the states on the survival level. Should this assumption be relaxed, we would, in general, have to investigate the features of the barrier also on the default level which effectively results into creating barrier trees for the default level. We henceforth consider the Assumption III.3.1 to be in effect. Let us now finally discuss the pricing of the barrier options itself. The pricing always involves two parts: the calculation of the DEPs contribution and the contribution of the survival states. Therefore

$$V = V_{0,0}^h + V(\kappa), \text{ with } V(\kappa) = \sum_k \kappa_{\cdot,\cdot,0,0}^{k,h}, \quad (\text{III.3.24})$$

where $V_{0,0}^h$ is the price contribution of the survival states,¹⁰⁹ $V(\kappa)$ represents the overall contribution of the defaulted states (DEPs) to the option price and $\kappa_{\cdot,\cdot,0,0}^{k,h}$ denotes the $(0,0)$ -value of the k th DEP adjusted for the barrier h .

Let us describe the **up-and-in** case first. To process the price contribution of the DEPs, which are seen as options, we need to create the 'activated' and 'inactivated' (barrier adjusted) tree for every DEP on the grid. A (i,j) -position of a DEP (i^*,j^*) with some index k on the activated tree will be denoted $\kappa_{i^*,j^*,i,j}^k$ and on the inactivated (barrier-adjusted) tree the point will be denoted $\kappa_{i^*,j^*,i,j}^{k,h}$ as already used above in (III.3.24). We build the activated tree of a DEP κ_{i^*,j^*} first. According to the Lemma III.3.2 we have at $i = i^*, j = j^*$

$$\kappa_{i^*,j^*,i^*,j^*}^k = (1 - \pi)^{i^*-1} \pi V_{i^*,j^*}^\delta.$$

For the time steps decreasing from $i = i^* - 1$ to $i^* = 0$ we apply the Lemma III.3.3. After the activated tree is created for every DEP, for every such a DEP we construct a paired inactivated tree. Thus, for DEP (i^*,j^*) , $0 \leq i \leq i^*$ compute

$$\kappa_{i^*,j^*,i,j}^{k,h} = \begin{cases} 0, & \text{if } i = i^* \\ \kappa_{i^*,j^*,i,j}^k & \text{if } i < i^* \text{ and } P_{i,j,S} \geq h, \\ \frac{1}{2}d_{i,j} \left(\kappa_{i^*,j^*,i+1,j+1}^{k,h} + \kappa_{i^*,j^*,i+1,j-1}^{k,h} \right) & \text{if } i < i^* \text{ and } P_{i,j,S} < h \end{cases} \quad (\text{III.3.25})$$

by iterating i backwards. Notice that at $i = i^*$ we have $\kappa_{i^*,j^*,i,j}^{k,h} = 0$ and this is because the option could not have been activated at time point $i = i^*$ by virtue of the Assumption III.3.1, since at $t = i^* \Delta_t$ the underlying P jumped to the default and thus the value $P_{i^*,j^*,S}$ where the activation is theoretically possible could not have been attained. The sum of all DEPs at $(0,0)$ $V(\kappa) = \sum_k \kappa_{\cdot,\cdot,0,0}^{k,h}$ gives the total contribution of the default states to the option price. It remains to define the contribution of the survival states $V_{0,0}^h$. This is done in a 'one-shot' procedure, that is without any need for a decomposition to smaller units. However, two trees V^A, V^h are still needed. Here V^A represents the 'activated' tree of the option prices on the survival level, and V^h is the inactivated (or barrier-adjusted) tree of the option prices on the survival level. Again, it is necessary to build the 'activated' tree V^A first and so we have

$$V_{i,j}^A = \begin{cases} (1 - \pi)^{N_T} \varphi_{N_T,j} & \text{if } i = N_T \\ \frac{1}{2}d_{i,j} \left(V_{i+1,j+1}^A + V_{i+1,j-1}^A \right) & \text{if } 0 \leq i < N_T \end{cases}.$$

¹⁰⁹see our remark above which states that the survival surface is considered to be a single, inseparable unit.

Notice that the survival probability $(1 - \pi)^{N_T}$ until the option's maturity $t = N_T \Delta_t$ had to be accounted for in the terminal time point N_T to specialize the tree only on the 'survival states' of the underlying. Having completed the activated tree V^A we compute the inactivated tree V^h so that the activation barrier h is taken into account. Hence

$$V_{i,j}^h = \begin{cases} V_{i,j}^A & \text{if } P_{i,j,S} \geq h \\ 0 & \text{if } i = N_T \text{ and } P_{N_T,j,S} < h, \\ \frac{1}{2}d_{i,j} \left(V_{i+1,j+1}^h + V_{i+1,j-1}^h \right) & \text{if } i < N_T \text{ and } P_{i,j,S} < h \end{cases} \quad (\text{III.3.26})$$

for i decreasing from $i = N_T$ to $i = 0$. Having the collection $\left\{ \kappa_{i^*,j^*,i,j}^{k,h} \right\}, \forall k$ of the inactivated (barrier-adjusted) trees of the quasi options reflecting the default states (eq. (III.3.25)) and the inactivated (barrier-adjusted) tree V^h reflecting the states on the survival level (eq. (III.3.26)), we can combine both these components and obtain the total barrier option price through the expression (III.3.24).

Pricing of **up-and-out** barrier options proceeds analogously but is simpler in the sense that the 'activation' trees are not needed. In brief, we firstly calculate the collection $\left\{ \kappa_{i^*,j^*,i,j}^{k,h} \right\}$ of the barrier-adjusted values of the DEPs. We have for i going from $i = i^*$ to $i = 0$

$$\kappa_{i^*,j^*,i,j}^{k,h} = \begin{cases} (1 - \pi)^{i^*-1} \pi V_{i^*,j^*}^\delta, & \text{if } i = i^* \\ 0 & \text{if } i < i^* \text{ and } P_{i,j,S} \geq h. \\ \frac{1}{2}d_{i,j} \left(\kappa_{i^*,j^*,i+1,j+1}^{k,h} + \kappa_{i^*,j^*,i+1,j-1}^{k,h} \right) & \text{if } i < i^* \text{ and } P_{i,j,S} < h \end{cases} \quad (\text{III.3.27})$$

The first expression at $i = i^*$ is again granted through the Assumption III.3.1 which ensures that the option could not have been knocked out at $t = i^* \Delta_t$ (but still it could have been knocked out in the prior time nodes). Then, we obtain the contribution of the survival states as

$$V_{i,j}^h = \begin{cases} 0 & \text{if } P_{i,j,S} \geq h \text{ and any } i \\ (1 - \pi)^{N_T} \varphi_{N_T,j} & \text{if } P_{i,j,S} < h \text{ and } i = N_T, \\ \frac{1}{2}d_{i,j} \left(V_{i+1,j+1}^h + V_{i+1,j-1}^h \right) & \text{if } P_{i,j,S} < h \text{ and } i < N_T \end{cases} \quad (\text{III.3.28})$$

Having both the collection $\left\{ \kappa_{i^*,j^*,i,j}^{k,h} \right\}$ of the barrier-adjusted default state contributors (eq. (III.3.27)) and the barrier-adjusted tree V^h reflecting the states on the survival level (eq. (III.3.28)), we can finally compute the total option price using the equation (III.3.24). We illustrate the method on a stylized hypothetical example and on a real-world one. The reason for inclusion of the hypothetical example is that it allows to graphically completely represent the trees that are part of the pricing problem.

III.3.7 Applications

As the methodology has been thoroughly discussed in the preceding sections, we illustrate its application on two worked-out examples. The first example is included rather for the pedagogical reasons. In that we provide a full analysis of the pricing scheme. We

intentionally select the set of parameters such that it does not produce large trees but rather simple ones which, however, capture all the features discussed above. The second numerical conduct is a real-world application on a sample data from the market. For completeness, we consider both the up-and-in and up-and-out barrier options in both cases.

Assume a corporate zero-bond P maturing at $s = 4$ with a recovery $\delta = 0.4$. It is assumed that the one-period default probability is $\pi = 0.05$. The put option written on the bond matures at $T = 3$, the strike is $K = 0.94$ and the barrier is $h = 0.88$. We assume that the time step has length $\Delta_t = 1$, the yield curve is flat, $t \mapsto r(0, t) = 0.06$, and the volatility of the log of the short rate is $\sigma = 0.5$. We have $N_T = 3, N_s = 4$. We start by constructing the trees $P_{i,j,S}, P_{i,j,D}$ as given by (III.3.9). These are shown in the Table 10. To construct the trees for P we needed also the tree of the one-period risk-free discount

j/i	0	1	2	3	j/i	0	1	2	3
3				0.8056	3				0.3322
2			0.7313		2			0.3107	
1		0.7071		0.9023	1		0.3093		0.3721
0	0.7039		0.8540		0	—		0.3628	
-1		0.8285		0.9439	-1		0.3624		0.3892
-2			0.9077		-2			0.3856	
-3				0.9602	-3				0.3960

Table 10: Bond prices $P_{i,j,S}$ on the survival level (left) and bond prices $P_{i,j,D}$ on the default level (right).

factors $d_{i,j}$ which we show in the Table 11 together with a tree of the option value on the defaulted surface V^δ . V^δ we constructed by (III.3.13). The barrier-adjusted trees V^h

j/i	0	1	2	3	j/i	0	1	2	3
3				0.8305	3				0.6078
2			0.8823		2			0.5187	
1		0.9184		0.9302	1		0.4830		0.5679
0	0.9434		0.9532		0	—		0.5332	
-1		0.9684		0.9731	-1		0.5185		0.5508
-2			0.9823		-2			0.5377	
-3				0.9899	-3				0.5440

Table 11: Tree of one-period BDT discount factors d (left) and a tree of option values V^δ at defaulted states of the underlying (right).

(both up-and-in and down-and-out) for the survival states are presented in the Table 12.

Finally, we shall provide also the (barrier-adjusted) trees for the DEPs. This would, however, occupy too much space when visualized and therefore we just summarize the contribution of the DEPs in the overviewing Table 13.

To calculate the total barrier option price for the up-and-in and up-and-out variants, we apply the formula (III.3.24) and read-off the respective values from the Table 12 and

j/i	0	1	2	3	j/i	0	1	2	3
3				0.0000	3				0.1152
2			0.0143		2			0.0508	
1		0.0136		0.0324	1		0.0233		0.0000
0	0.0100		0.0154		0	0.0110		0.0000	
-1		0.0075		0.0000	-1		0.0000		0.0000
-2			0.0000		-2			0.0000	
-3				0.0000	-3				0.0000

Table 12: Trees of the barrier-adjusted option prices V^h on the survival level. The left tree is for the up-and-in barrier, the right tree is for the up-and-out barrier.

k	i^*	j^*	κ_{i^*,j^*,i^*,j^*}	$\kappa_{i^*,j^*,0,0}$	$\kappa_{i^*,j^*,0,0}^h$ (in)	$\kappa_{i^*,j^*,0,0}^h$ (out)
1	1	1	0.024150735	0.011391856	0.000000000	0.011391856
2	1	-1	0.025924159	0.012228377	0.000000000	0.012228377
3	2	2	0.024636163	0.005336468	0.000000000	0.005336468
4	2	0	0.025325547	0.011269823	0.000000000	0.011269823
5	2	-2	0.025539670	0.005832930	0.000000000	0.005832930
6	3	3	0.027426981	0.002620768	0.000000000	0.002620768
7	3	1	0.025628085	0.007884271	0.000000000	0.007884271
8	3	-1	0.024852661	0.008058606	0.002787669	0.005270937
9	3	-3	0.024549066	0.002753615	0.002753615	0.000000000

Table 13: Summary table of all 9 DEPs. The fourth column from the right κ_{i^*,j^*,i^*,j^*} holds the value of the DEPs produced by the Lemma III.3.2. The third column from the right $\kappa_{i^*,j^*,0,0}$ shows the present value of the DEPs when no barrier is in place (given by the Lemma III.3.3 and (III.3.23)). The barrier-adjusted present values of the DEPs are in the second (up-and-in) and first (up-and-out) column from the right.

the Table 13. This gives us

$$\begin{aligned} V_{\text{up-and-in}} &= 0.00995535179 + 0.00554128394 = 0.01549663573 \\ V_{\text{up-and-out}} &= 0.01101210055 + 0.06183543079 = 0.07284753134. \end{aligned}$$

It is critical that the model does not allow arbitrage by trading in these barrier options and European options (see the Section I.1.4 for more details). The arbitrage will be prevented if the parity equation III.3.21 is satisfied. Computing the vanilla plain European option price gives 0.0883441671 and the sum of the barrier options also gives $V_{\text{up-and-in}} + V_{\text{up-and-out}} = 0.0883441671$. This ensures that the tree model is free of arbitrage. Notice that the model is arbitrage-free even in the case of every single individual DEP. For example, for the DEP $k = 8$ (Table 13), the 'in' price of that point is 0.002787669, the 'out' price is 0.005270937. In total the (in+out) price is 0.008058606 which is equal to the value of the point if there were no barriers in effect.

The real-world example is worked-out similarly. We, however, suppress any graphical representation of the trees because it would consume too much space. For the analysis, we have selected a bond issued by Dell (ISIN US24702RAJ05). The maturity of the bond is June 15, 2019, the bond pays coupons at rate 5.875 % semiannually. The current price of the bond, as of the valuation date May 15, 2016 is 104.25. From (Ou et al., 2016), the expected recovery of the bond is $\delta = 0.379$ (Senior Unsecured).¹¹⁰ The option written on this bond matures at $T = 1$, has strike $K = 1.045$ and is subject to a barrier $h = 1.05$ (again, we assume both up-and-in and down-and-out variants). First of all, the interest-rate model (BDT) needs to be calibrated. For that we have chosen quotes of ATM caps of all available tenors (1Y - 30Y). The calibration fit is shown in Figure 25, where the points assigned to the lower premiums typically represent cap quotes for the shorter tenors. It is thus easy to see that the fitted BDT model was able to capture quite well the quotes for the shorter tenors which is sufficient for the selected pricing problem.

¹¹⁰the average recovery of the Sr. Unsecured bonds is a surprisingly stable quantity in time. Typically, its value fluctuates between $\delta = 0.3$ and $\delta = 0.4$ in the Moody's Annual Default Study report.

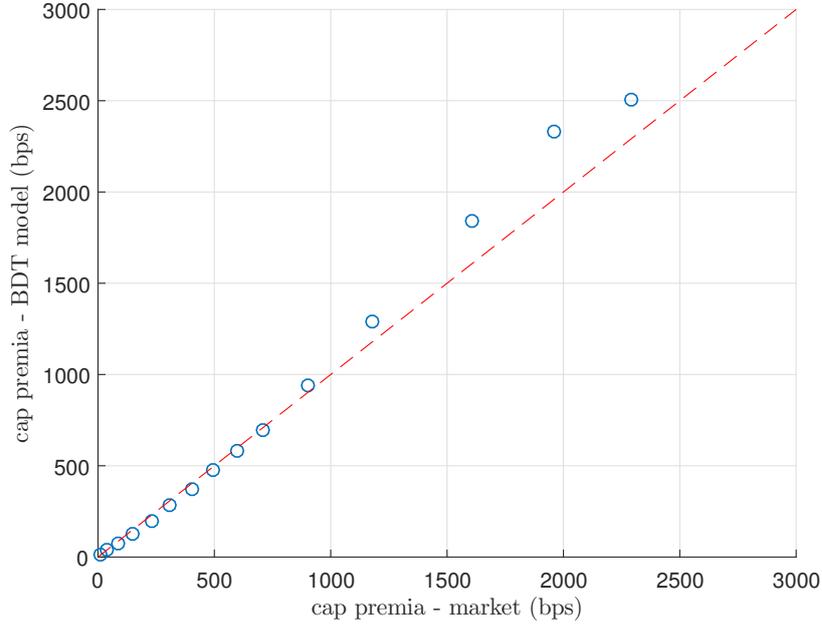


Figure 25: BDT model prices vs market prices of ATM caps as of May 15, 2016. The dashed red line is a perfect fit line.

The optimal BDT parameter resulting from this calibration was $\sigma = 0.5893$. Having the *only* necessary interest-rate parameter, we need to estimate the one-period default probability π connected with some period of length $\Delta_t = 0.125$ which we have chosen to ensure a fineness of the tree.¹¹¹ Optimal π can be found by matching the observed market price of the defaultable bond (104.25) with the model bond price P , by iterating π in (III.3.9) (zero-bond), or in (III.3.10) (coupon-bearing bond). In our case the corporate bond is coupon-bearing and so the optimal π found by (III.3.10) turns out to be $\pi = 0.01684$. After this is done, we can construct all the necessary trees to price the barrier option, similarly to the previous simplified case. We find

$$\begin{aligned}
 V_{\text{up-and-in}} &= 0.001972233313282 + 0.068235337753582 = 0.070207571066864 \\
 V_{\text{up-and-out}} &= 0 + 0.016824174668240 = 0.016824174668240.
 \end{aligned}$$

The price of the European option with (no barrier) on this bond is equal to $V = 0.08703175$. It is easy to see that it is equal to the sum $V_{\text{up-and-in}} + V_{\text{up-and-out}}$ and so, again, the parity (III.3.21) holds. This ensures the model is free of arbitrage.

III.3.8 Numerical validation against Monte Carlo and PDE

The computationally undemanding model we have proposed can and in fact should be benchmarked against some general approaches to verify that the results are correct. The

¹¹¹this is not perfectly compatible with the 3M tenor, implying $\Delta_t = 0.25$ which was used for estimation of the 'short-rate' parameter σ but we can regard 3M ($\Delta_t = 0.25$) rate and 6W ($\Delta_t = 0.125$) rate to be reasonably close and both to be proxies for the hypothetical short-rate.

numerical verification, however, is a *surprisingly* difficult issue and requires a combination of MC both and the PDE methods. The main issue is that P can not be computed using PDE because in the PDE (reduced-form models; see the Section III.2) we do not know if the default has already occurred or not, it is a backward-operating method. However, the P -equation (III.3.2) requires the knowledge of the default or the survival state. This enforces to use MC. Yet B as a component of P , on the other hand can not be computed by MC because we would have a simulation in simulation when pricing the bond on which the option is written.

Assume just a European option on a defaultable zero bond with no barriers involved. Then, to verify the model, we have to check the \mathbb{Q} -expected option payoff at T and discount it back to time 0. If we had a closed form formula for $P(T, s)$ it would not be a difficult task. Although we have (III.3.2), we still have no formula for $B(T, s, r)$ because it is BDT-driven and the bond formula (see the Appendix (A.2)) in BDT has no closed-form. Hence, $B(T, s, r)$ has to be computed numerically using the SDE and the formula

$$d \ln r(t) = \theta(t)dt + \sigma dW^{\mathbb{Q}}(t), \text{ and } B(T, s, r) = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_T^s r(u)du \right) \middle| r(T) = r \right],$$

respectively. This can theoretically be handled by either MC or PDE. There are two main concerns, however. (i) we don't know $\theta(t)$, (ii) if $r(T) = r$ is about to be computed by MC, then $B(T, s, r)$ computed by MC is a 'simulation in simulation'. This would suggest using the PDE method everywhere. However, the PDE is not suitable for computation of $P(T, s, r)$ because it can not take into account the possible default occurring at a prior time than T . In turn, MC and PDE methods have to be combined so that PDE handles the risk-free bond B and MC models the default risk and the option price. The issue (i) can be overcome by a PDE for the bond price in BDT, which is

$$\frac{\partial B}{\partial t} + \theta(t) \frac{\partial B}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 B}{\partial x^2} - e^x B = 0, \text{ for } x = \ln r. \quad (\text{III.3.29})$$

Using this PDE we have to 'bootstrap' $\theta(t)$ by matching the bond prices (discount) curve. This is still a big issue because $\theta(t)$ tends to explode if suitable restrictions on $\theta(t)$ and tolerance to the fit to the discount curve are not imposed.¹¹² Notice that σ can be taken from the calibrated tree model because it is a fixed parameter same for the tree and the continuous SDE. After we have performed this numerical search for $\theta(t)$, it can be seen in Figure 26 together with the obtained discount curve fit. Now, once having $\theta(t)$ we can use it either in MC or the PDE method. To prevent the 'simulation in simulation', when pricing the option expiring at T on s -maturity defaultable zero bond P we resorted to taking the following steps:

1. Simulate r up to time T . In the simulation use θ estimated above.
2. Simulate indicator of default $\mathbf{1}_{\{\tau^* \leq T\}}$. This ensures that at the option expiration T we are able to determine whether default has occurred or not.

¹¹²similar issues were reported by (Hogan and Weintraub, 1993).

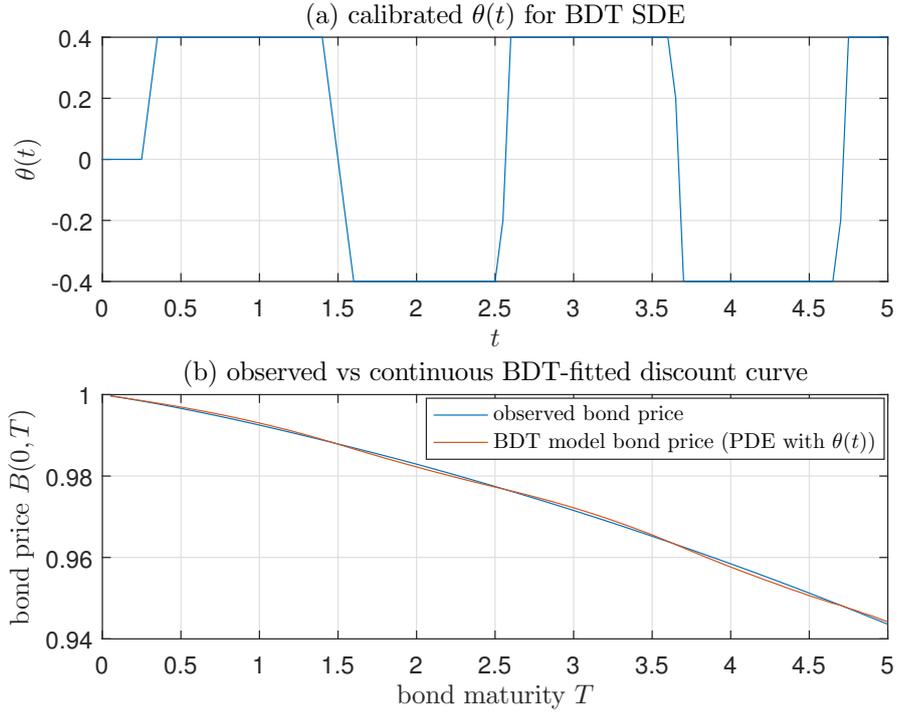


Figure 26: (a) shows the numerically obtained function $\theta(t)$ in continuous BDT. (b) shows the quality of continuous BDT bond prices fit to the observed discount curve.

3. For every $t \leq T$, compute the s zero-bond $B(t, s, r)$ ¹¹³ price using the PDE (III.3.29), where r is the simulated rate at time t . This also solves the issue (ii) above.
4. Compute the price of the defaultable zero bond at expiration T using (III.3.2), i.e. $P(T, s, r) = \mathbf{1}_{\{\tau^* \leq T\}} \delta B(T, s, r) + \mathbf{1}_{\{\tau^* > T\}} B(T, s, r) (\mathbb{Q}[T < \tau^* \leq s | \tau^* > T] (\delta - 1) + 1)$.
5. Calculate the payoff of the option V on the defaultable zero $P(T, s, r)$ at T and discount using the previously simulated rates r back to time 0.
6. As usual for MC, average the outcomes of the step 5. to get the value of the option.

If the barriers are involved then at every $t \leq T$ one needs to compute $P(t, s, r)$ (here we use the whole set of the default free s -zeros $t \in [0, T] \mapsto B(t, s, r)$ generated in the step 3.) and check whether the barrier was or was not hit in the simulated scenario. This information then needs to be incorporated into the effective payoff.

We have selected the following parameters for the test: $\sigma = 0.5893$, discount curve as shown in the Figure 26(b), $\Delta_t = 0.05$, $\pi = 0.00262$ (this is related to the intensity $\gamma = 0.0525$), $\delta = 0.3790$, $K = 0.90$, put option expiring at $T = 1$ on a defaultable zero bond maturing at $s = 3$. Then, we have run the complicated computation through MC combined with PDE and our implementation of the two-level defaultable tree. The computations we conducted for various levels of the barrier ranging from 0.85 to 0.95 for

¹¹³for the European options it is actually sufficient to compute just $B(T, s, r)$, yet the whole set of bond prices $t \in [0, T] \mapsto B(t, s, r)$ is necessary for the barrier options.

both the up-and-out and the up-and-in barrier options. The results for both methods are shown in the Table 14. The results confirm the barrier tree model that we proposed is consistent with the 'standard' approaches. However, one computation took only 0.1s with the developed tree methodology while it took nearly 15 seconds with the combined method MC + PDE (80 000 simulations).

barrier h	up-and-out		up-and-in	
	defaultable tree	MC and PDE	defaultable tree	MC and PDE
0.85	0.0000	0.0000	0.0277	0.0277
0.86	0.0000	0.0000	0.0277	0.0280
0.87	0.0000	0.0000	0.0277	0.0276
0.88	0.0000	0.0000	0.0277	0.0276
0.89	0.0072	0.0064	0.0206	0.0212
0.90	0.0147	0.0141	0.0131	0.0137
0.91	0.0206	0.0206	0.0072	0.0071
0.92	0.0255	0.0253	0.0022	0.0021
0.93	0.0277	0.0279	0.0001	0.0001
0.94	0.0277	0.0281	0.0000	0.0000
0.95	0.0277	0.0277	0.0000	0.0000

Table 14: Comparison of the computed barrier option prices on a defaultable bond achieved with the developed methodology (defaultable tree) and with a combination of the 'standard methods' (Monte Carlo and PDE).

Part IV

Portfolio problems

In the Part II and the Part III, we have discussed two main credit model classes: the firm-value models and the reduced-form models, yet we were mostly focusing on the single-entity level of credit risk and we have not discussed any portfolio problems. This Part IV thus makes the exposition complete in the sense that we add the portfolio-level problems to the topics (firm-value and reduced-form models) covered earlier. This arrangement is useful because the portfolio-level problems encompass subproblems related to both the firm-value and the reduced-form models.

Leaving aside the theory discussed earlier, we start this Part IV directly with a CDO modeling problem discussed in the following Section IV.1, where the main objective is to find an optimal approach for matching the market-quoted prices such that non-standard products can be priced after having a good initial fit (calibration). The major role in this section played by a double t copula model, whose semi-analytical properties we revealed in (Kolman, 2014a). We benchmark the model against the standard market model (Gaussian copula) and some other (NIG model, Archimedean copulas) and find that the double t model best describes the CDO credit market across various periods in time.

Finally in the Section IV.2, based on (Kolman, 2014a), we present a fairly general semi-analytical portfolio modeling framework that is compatible with the regulatory approaches and can easily be adjusted to capture the credit market phenomena the approximating regulatory formulas can not (e.g. stochastic recoveries). This is useful in the sense that a bank does not need to calculate any further 'adjustments' to patch the weak points in the regulatory formula. Also, the framework can work with the double t copula and we find that after using this specification, the model produces results close to that of (Witzany, 2013), who finds the standard regulatory formula to underestimate the true credit risk taken by the financial institutions.

IV.1 Synthetic CDO modeling

CDOs play an important role in today's credit markets as they allow the investors to take positions in a basket of credits which is structured into tranches ranging from junior (riskier) tranches up to senior (safe) tranches. Hence, CDOs are attracting both the risky and the risk-averse investors. In this text, we focus on an important subclass of CDOs, which are the synthetic CDOs referencing the names in the iTraxx index. We will henceforth use the term 'CDOs' for these specific single tranche synthetic CDOs. For a complete overview of CDOs and other structured credit products from the non-technical point of view, we refer the reader to (Rajan, 2007). Some other practical information can also be found in (Servigny and Jobst, 2007).

The main engineering question is how to optimally price the CDOs and mainly how to price the non-standard (or *bespoke* tranches). In the business practice, there are two main families of approaches: the bottom-up and the top-down approaches. The bottom-up approaches are based on modeling single defaults and from these the portfolio loss is

obtained by linking the defaults together typically by using copulas.¹¹⁴ The top-down approaches take the portfolio of credits as a whole and model the aggregated portfolio losses directly without inspecting the individual credits. For example, (Choroś-Tomczyk et al., 2016) model the whole 'correlation surface' instead of going into the detail of the individual credits. Alternatively, (Cont and Minca, 2013) model the default intensity of the whole portfolio. In this part we investigate the bottom-up approaches in the view of various copulas that determine the interconnectedness of the credits in the portfolio. Our main hypothesis is whether the double t copula outperforms the other copulas in CDO prices modeling. We also provide a semi-analytical representation of the double t copula which might be more challenging to deal with than the other copulas. Before we outline the CDO mechanics and the pricing logic, we firstly briefly discuss the common market practice.

IV.1.1 The current CDO market practice and its imperfections

The nowadays model standard for the CDO market is still the Gaussian copula proposed by (Li, 1999) although there are many related difficulties and imperfections. Historically, it was believed that the correlations in the CDO models reflect the empirical correlations. Under this assumption, when correlations are the objective inputs to the model, tranche prices (premiums) come out as the outcome. The main related problem is that if a trader wants the tranche to have a different price then what would correspond to the empirical correlation, this can only be achieved by quoting a different correlation than the empirical one because there are no other inputs he can change in order to affect the price. This rules out the use of the empirical correlations and forces the traders to use artificial, so called *implied* correlations in order to generate the prices that they want. This is conceptually similar to the use of the implied volatilities in option pricing. Secondly, the CDO portfolio is typically partitioned into several tranches and traders might want to set a special price to every tranche. This again implies that every tranche has to be priced with a different correlation. Finally, the tranche premium might not be a strictly monotonous function of the correlation. This causes the prices of the tranches for chosen correlation to be non-unique. To circumvent the non-uniqueness of the 'correlation-price' relationship, the concept of the base correlation (Lee McGinty, 2004) has been developed. The use of the base correlations, however, has its other deficiencies¹¹⁵ yet it remained a standard that the base correlations are quoted quantities.

For other common pitfalls related to Gaussian copula for the CDO modeling we refer the reader to (Brigo, 2010) who provides an exhaustive list of issues related with the Gaussian copula. The major drawback of the Gaussian copula that we find is that it is difficult to be used for pricing of the bespoke tranches because a single parameter (correlation) does not fit all the tranche spreads. It is thus a very desired feature that the model can produce unique prices of tranches with respect to trivial changes in the

¹¹⁴the theoretical foundations of 'copulas' is not covered here. Our thoughts are mostly based on the exposition provided in (McNeil et al., 2005) and (Embrechts et al., 2001).

¹¹⁵the main problem with the base correlation is the inconsistency across the whole capital structure. For example the 0%-3% tranche is priced via a different base correlation than the 0%-22% tranche although the 0%-3% tranche is actually a subtranche of the 0%-22% tranche.

parameters. Our main motivation for the analysis is thus to find a model that with a single set of parameters provides a close match to market prices. This in turn is then almost surely an appropriate model for pricing bespoke tranches. Before we discuss further details, it is useful to revisit the mechanics of a CDO deal.

IV.1.2 The mechanics of a CDO deal

A CDO tranche deal involves two counterparties: long CDO (protection seller¹¹⁶) and short CDO (protection buyer). The protection seller is typically the investor who buys the default risk and receives a coupon (premium) from the protection buyer. Commonly, the protection buyer who pays the coupons is some investment bank, usually referred to as 'the dealer'. The contracted deal is related to a credit basket of n names¹¹⁷ which is structured into tranches ranging from the equity (junior) tranche up to the super senior tranche with a simple waterfall structure. The dealer then provides market making (quotes) for these tranche deals. The most active CDO market is that of the CDOs on iTraxx credits (European region) or CDX.NA credits (North American region). The composition of these baskets is, however, managed by an independent company Markit.¹¹⁸ The credits in the baskets are subject to a so called 'semiannual roll' which involves adjustments in the credit selection. Credits that no longer satisfy the selection conditions are replaced by new credits. However, the old (pre-roll) series still continues to run and thus deals on this series can still be traded. The importance of Markit also resides in that it defines the attachment and detachment points of the tranches. In general the whole CDO standardization and transparency maintained by Markit significantly increases the liquidity of the product. In practice, however, the dealers typically use the standardized tranches as benchmarks and the real deals with investors are customized into to bespoke deals.

From the cash-flow point of view, it is common for the junior tranche deals that an upfront payment u is settled between the protection seller and the buyer at initiation of the CDO deal. *The upfront cash-flow is seen from the protection seller's (investor's) perspective* and so if $u > 0$ then an upfront payment is paid by the protection buyer to the seller and if $u < 0$ then an upfront payment is paid by the protection seller to the buyer.¹¹⁹ This is shown in the Figure 27. During the life of the CDO deal the protection buyer pays coupons on the outstanding tranche principal to the protection seller and the protection seller pays the losses of the tranche to the protection buyer as portrayed in the Figure 28. In the context of the CDO trading, the coupon rate, or spread, however, has to be correctly understood. There are two types of spreads (or coupons), in general. Firstly, there is a theoretical fair spread s and secondly there is a fixed running coupon s^* . The

¹¹⁶here, the term *long* CDO related with the protection *seller* sounds counterintuitive because 'long' usually stands for the 'buyer'. It is thus more natural to consider long CDO as a risk buyer, rather than a protection seller.

¹¹⁷in the case of the main iTraxx or CDX.NA $n = 125$.

¹¹⁸although Markit is independent from any market participants, it still selects the eligible credits according to a dealer poll.

¹¹⁹for the senior tranches there is usually no upfront payment, though.

theoretical fair spread is a trivial quote – it is such a coupon rate which makes the present values of the premium and the protection leg equal so that no up-front is settled at the inception. The fixed running spread s^* is an enforced rate under which the deal actually runs. This clearly means that the PV of the premium and the protection leg under s^* will not be equal and this difference is allocated into the upfront u . It is customary that for the junior tranches the fixed running spread + upfront mechanics is used while the senior tranches actually run under the fair spread with no upfront payment.

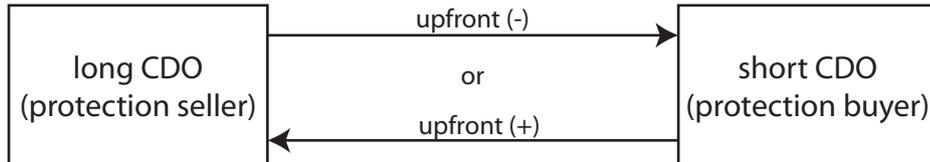


Figure 27: Initiation of a standardized synthetic CDO deal.

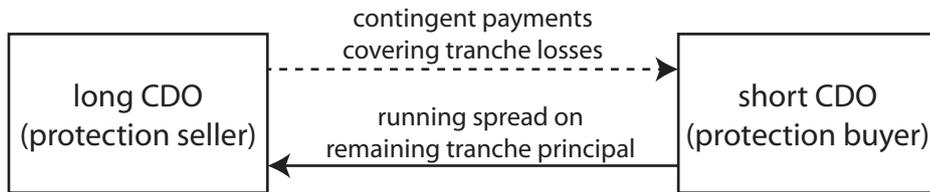


Figure 28: Standardized synthetic CDO deal during its life.

To make this more practical, some sample dealer quotes are then summarized in the Table 15. If, for example, an investor sells protection on the 5Y 3%-6% tranche, he will receive $u = 0.03945$ of the deal notional upfront and then the deal will physically run under $s^* = 0.05$. If he instead entered the same tranche of a 3Y deal, he would *pay* 0.02180 ($u = -0.02180$) to the protection buyer. If the investor wanted to sell the protection on e.g. the 5Y 9%-12% tranche, he would be receiving $s = 0.0404$ and there would be no upfront payment at the deal inception. Notice the 'Base correlation' quotes next to the 'Spread/Upfront' quotes. The base correlation is so important for the dealers that it is a part of the quotes.

IV.1.3 General pricing framework

Now that the CDO mechanics is laid out, we will outline the general pricing methodology. The idea is essentially similar to that of the CDS pricing,¹²⁰ in that the fair spread s is such a coupon rate which equalizes the present value of the expected incurred losses with the present value of the expected coupons received for the protection. Let a, d denote the tranche *attachment* and the *detachment* point, respectively. Then the fair spread s on a CDO tranche $[a, d]$ satisfies

$$s\mathcal{A}_{[a,d]} + s\mathcal{B}_{[a,d]} - \mathcal{C}_{[a,d]} = 0, \quad (\text{IV.1.1})$$

¹²⁰notice the similarity with CDS, as the synthetic CDO is nothing but a position in a basket of CDS.

Tranche	Spread/Upfront		Base correlation	
	3 year	5 year	3 year	5 year
0-3 (%) [*]	16.600	35.490	59.322	55.802
3-6 (%) [*]	-2.180	3.945	62.666	62.398
6-9 (%) ^{**}	221.000	404.000	66.730	65.967
9-12 (%) ^{***}	132.440	260.000	70.973	70.193
12-22 (%) ^{***}	71.750	137.250	82.614	83.625
22-100 (%) ^{****}	21.625	43.690	-	-

Table 15: Bloomberg quotes as of June 28, 2013 for tranches of iTraxx EUR, series 19. Quotes denoted by (*) are quoted in (%) of notional upfront with running spread 500 bps, (**) is quoted in bps with running spread 300 bps, (***) is quoted in bps with running spread 100 bps, (****) is a bps quote with running spread 25 bps. Correlations are in (%). Overall index quotes were 73.742 bps for 3-year maturity and 117.207 bps for 5-year maturity. The 3Y deal matures on Jun 20, 2016 and the 5Y deal matures on Jun 20, 2018.

where $s\mathcal{A}_{[a,d]}$ is a present value of the coupons payments, $s\mathcal{B}_{[a,d]}$ is the present value of the accrual coupon component¹²¹ and $\mathcal{C}_{[a,d]}$ is the present value of the losses of the tranche $[a, d]$. This means that the fair spread s equals

$$s = \frac{\mathcal{C}_{[a,d]}}{\mathcal{A}_{[a,d]} + \mathcal{B}_{[a,d]}}.$$

In the previous Section IV.1.2 we have discussed that there is a difference between the fair spread s and the fixed running spread s^* and hence it is clear that if the deal runs under s^* instead of s the parity (IV.1.1) would, in general, not hold (since s is replaced by $s^* \neq s$ in that equation). When s is replaced by s^* in (IV.1.1), then in order to make the expression an equality, an extra term u needs to be introduced, whose quantity is again seen from the protection seller's perspective and so

$$s^* \mathcal{A}_{[a,d]} + s^* \mathcal{B}_{[a,d]} - \mathcal{C}_{[a,d]} + u = 0 \Rightarrow u = \mathcal{C}_{[a,d]} - s^* (\mathcal{A}_{[a,d]} + \mathcal{B}_{[a,d]}).$$

Notice here that if $s^* < s$ then, the protection seller receives less than would be fair and he thus has to receive $u > 0$ upfront to be compensated for the small coupon. If, conversely, $s^* > s$, then the protection seller receives more than what would be fair and so $u < 0$ which means he has to compensate the protection buyer upfront. It is also very useful to relate the fair spread s to the fixed running spread s^* using the last equation when substituting for $(\mathcal{A}_{[a,d]} + \mathcal{B}_{[a,d]})$ from (IV.1.1). This gives after some trivial simplifications

$$u = \mathcal{C}_{[a,d]} \left(1 - \frac{s^*}{s} \right).$$

IV.1.4 The terms $\mathcal{A}, \mathcal{B}, \mathcal{C}$ in a bottom-up framework

In the previous section the fundamental relations that must hold for a CDO tranche price were set up. These relations are general and must hold all the time regardless of the model

¹²¹the price contribution of this component is rather marginal. For this reason, sometimes it is assumed away. See for example (Fusai and Roncoroni, 2008, Section 22.2.1) who do not take the accrual spread into account in the case of basket default swaps.

being used. As we want to apply the 'bottom-up' framework it is necessary to specialize the components $\mathcal{A}, \mathcal{B}, \mathcal{C}$ for the bottom-up setup. The main idea behind the bottom-up single factor framework that we are about to use is that there exists a general economic factor X affecting the health of all the credits in the basket. Conditional on state of this factor, the credits are assumed to be independent.

Let $\mathcal{A}_{[a,d]}|x$ denote the amount of expected coupon payments (in the present value) given the common factor X is at state x . Since the spread is only paid on the remaining tranche principal, we need to take this into account. Let $\mathbb{E}^{\mathbb{Q}}[E(t)|x] = \mathbb{E}^{\mathbb{Q}}[E(t)|X=x]$ denote the expected tranche principal at time t as seen from the time zero, given $X=x$. Because the premium is paid in discretely at times $t_1 > 0, t_2, \dots, t_N$ such that $\Delta \equiv \Delta(t_i, t_{i-1})$, we have

$$(\mathcal{A}_{[a,d]}|x) = \Delta \sum_{i=1}^N \mathbb{E}^{\mathbb{Q}}[E(t_i)|x] B(0, t_i).$$

As the real-world defaults normally occur not only at the coupon dates but between the coupons and the accrual part is also paid to the protection seller it is a good practice to account for this accrual part in the computations and we have

$$\begin{aligned} (\mathcal{B}_{[a,d]}|x) &= \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \Delta(t_{i-1}, t) B(0, t) d\left(-\mathbb{E}^{\mathbb{Q}}[E(t)|x]\right) \\ &\approx \frac{1}{2} \Delta \sum_{i=1}^N B\left(0, \frac{t_{i-1} + t_i}{2}\right) \left(\mathbb{E}^{\mathbb{Q}}[E(t_{i-1})|x] - \mathbb{E}^{\mathbb{Q}}[E(t_i)|x]\right), \end{aligned}$$

where the assumption (in the approximation) was adopted that the default approximately occurs in the midpoint of two consecutive coupon dates. Notice the integration goes over the differential of the *negative* expected exposure in time. This is because the expected exposure is, ceteris paribus, a decreasing quantity in time and so the if the minus sign were not included the differential would be negative and the $(\mathcal{B}_{[a,d]}|x)$ term would be negative too which is impossible by definition. The last but critical component is the present value of the expected tranche losses. We have

$$\begin{aligned} (\mathcal{C}_{[a,d]}|x) &= \int_0^T B(0, t) d\left(-\mathbb{E}^{\mathbb{Q}}[E(t)|x]\right) \\ &\approx \sum_{i=1}^N B\left(0, \frac{t_{i-1} + t_i}{2}\right) \left(\mathbb{E}^{\mathbb{Q}}[E(t_{i-1})|x] - \mathbb{E}^{\mathbb{Q}}[E(t_i)|x]\right), \end{aligned}$$

where the same discretization logic as for $(\mathcal{B}_{[a,d]}|x)$ to transform the integral into a simple approximating sum is used.

Once the conditional values are defined, we integrate out the factor X by integrating over the density of X and obtain the unconditional values $\mathcal{A}_{[a,d]}, \mathcal{B}_{[a,d]}, \mathcal{C}_{[a,d]}$ needed for the the spread formula (IV.1.1). Thus, denoting D the *support*¹²² of f_X , the unconditional

¹²²the support D of a function f is defined by $D = \text{supp}(f) = \{x : f(x) \neq 0\}$.

values are computed simply as

$$\begin{aligned}\mathcal{A}_{[a,d]} &= \int_D (\mathcal{A}_{[a,d]}|x) f_X(x) dx, \mathcal{B}_{[a,d]} = \int_D (\mathcal{B}_{[a,d]}|x) f_X(x) dx, \\ \mathcal{C}_{[a,d]} &= \int_D (\mathcal{C}_{[a,d]}|x) f_X(x) dx.\end{aligned}$$

To end this section, allow us to briefly comment the technical profile of the formulas. It is easy to see that the (conditional) expected tranche principal $\mathbb{E}^{\mathbb{Q}}[E(t_i)|x]$ plays a central role in the CDO pricing because it essentially determines the values of all the central pricing components. It is also computationally the most demanding quantity to be obtained. A good implementation of any single-factor model should resort to use only a limited number of calls of $\mathbb{E}^{\mathbb{Q}}[E(t_i)|x]$. We recommend to precompute $\mathbb{E}^{\mathbb{Q}}[E(t_i)|x]$ for all t_i and all x in question and then subsequently use these pre-cached values in evaluation of all the components above.

IV.1.5 The expected tranche principal

The essential component of the CDO modeling is the tranche principal E as coupons are paid with respect to the amount E and also the losses on the tranche are computed with respect to changes in E . The objective of this part is to provide techniques of construction of the $X = x$ -conditional expected tranche principal $\mathbb{E}^{\mathbb{Q}}[E(t)|x]$. Here, we mostly follow (Hull, 2008). Some authors (for example (Kalemanova et al., 2005), (Kane, 2008) or (Schlosser, 2011)), however, prefer to use the expected tranche loss rather than the outstanding principal, yet both approaches are equivalent.

To start, assume there is n credits in the basket and assume the portfolio is homogeneous in the sense that all the credits have the same credit risk profile (default probability and the recovery δ). One default thus causes a loss of delta $(1 - \delta)$ in the whole basket of n credits, whose notional is n (each credit has a notional 1). If k defaults occur, the total loss is $k(1 - \delta)$ and when expressed relatively with respect to the whole notional n , the credit loss is $z = \frac{k(1-\delta)}{n}$. This means the tranche in consideration starts to be affected when $z \geq a$ and is completely wiped out once $z \geq d$. Hence, there are some critical integer values of the number of defaults k which we denote k_a, k_d where the tranche first is affected and is completely wiped out, respectively. If we set $a = \frac{k(1-\delta)}{n}$ and solve for k we get $k = \frac{an}{1-\delta}$ yet this can, in general, be a real number (not necessarily an integer) and we have to round up to the nearest greater integer. The same holds for the the tranche

detachment point d .¹²³ Hence¹²⁴

$$k_a = \left\lceil \frac{an}{1-\delta} \right\rceil, k_d = \left\lceil \frac{dn}{1-\delta} \right\rceil.$$

Now that the number of defaults that affect the tranche is specified, we will quantify how much of the tranche principal is wiped out if $k, k_a \leq k < k_d$ defaults occur. It is obvious that the size or length of the tranche is simply $(d-a)$. Recall that k defaults cause $\frac{k(1-\delta)}{n}$ relative loss in the whole portfolio. Given $k, k_a \leq k < k_d$, the tranche has lost $\frac{k(1-\delta)}{n} - a$ and this quantity divided by the tranche size gives the tranche's relative loss is $\frac{\frac{k(1-\delta)}{n} - a}{d-a}$. We can thus easily express the remaining tranche principal $E(t)$ in terms of the number of losses k as

$$E(t) = \begin{cases} 1, & \text{if } k < k_a \\ 1 - \frac{\frac{k(1-\delta)}{n} - a}{d-a}, & \text{if } k_a \leq k < k_d \\ 0, & \text{if } k \geq k_d \end{cases}$$

It remains to define the *expected* tranche principal which essentially means using the last expression and enrich it for the \mathbb{Q} -probability of k defaults to occur, gives some value of X . This yields¹²⁵

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[E(t)|x] &= \sum_{k=0}^{k_a-1} \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{k \text{ defaults until } t\}}|x] \\ &+ \sum_{k=k_a}^{k_d-1} \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{k \text{ defaults until } t\}}|x] \left(1 - \frac{\frac{k(1-\delta)}{n} - a}{d-a}\right). \end{aligned}$$

To simplify the notation, let $\mathbb{Q}(t, k|x) := \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{k \text{ defaults until } t\}}|x]$ and we obtain a compact expression

$$\mathbb{E}^{\mathbb{Q}}[E(t)|x] = \sum_{k=0}^{k_a-1} \mathbb{Q}(t, k|x) + \sum_{k=k_a}^{k_d-1} \mathbb{Q}(t, k|x) \left(1 - \frac{\frac{k(1-\delta)}{n} - a}{d-a}\right).$$

It will be shown in the subsequent sections that the probabilities of any credit to default which underlie $\mathbb{Q}(t, k|x)$ represent the core of the CDO modeling.

In this section we further propose a numerically efficient way of computing the probability vector $\mathbb{Q}(t, k|x), k = 0, \dots, n$ which outperforms the classical 'binomial' method

¹²³for example, if we had the tranche $[a, d] = [3\%, 6\%]$ in a basket of $n = 125$ credits, each credit with recovery $\delta = 0.4$, then $\frac{an}{1-\delta} = 6.25$, $\frac{dn}{1-\delta} = 12.5$. If 6 defaults occur, there is no tranche loss because the number of defaults is below 6.25. So, the tranche gets first affected at $k_a = 7$ defaults. Similarly, if there are 12 defaults, the tranche is still not wiped out completely because it would be if there were 12.5 defaults, yet this can not empirically occur. In turn the tranche gets first completely wiped out once there are $k_d = 13$ defaults.

¹²⁴the operation $\lceil x \rceil$ denotes the *ceiling* function applied to x . That is $\lceil x \rceil$ is the smallest integer greater than or equal to x .

¹²⁵notice there could have been included a third summation $\sum_{k=k_d}^n \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{\{k \text{ defaults until } t\}}|x] \times 0$ for completeness, yet this has no numerical effect on the expression $\mathbb{E}^{\mathbb{Q}}[E(t)|x]$.

in speed while providing equally accurate results. Assume for now that we have somehow computed the X -conditional probability $Q(t|x)$ of any name of the n -name basket to be in default by time t and further assume that conditionally on $X = x$, all the credits are *independent*, as usual. Denote $\mathbb{Q}(t, \mathbf{k}|x)$ the $(n + 1)$ -dimensional vector $\mathbb{Q}(t, \mathbf{k}|x) = [\mathbb{Q}(t, 0|x), \mathbb{Q}(t, 1|x), \dots, \mathbb{Q}(t, n|x)]$. If we use the standard binomial approach for the computation of the probability $\mathbb{Q}(t, k|x)$ of observing exactly k defaults by time t , then to fill $\mathbb{Q}(t, \mathbf{k}|x)$ it is necessary to call

$$\mathbb{Q}(t, k|x) = \binom{n}{k} Q(t|x)^k (1 - Q(t|x))^{n-k}, k = 0, 1, \dots, n,$$

which can eventually consume a lot of resources as the binomial coefficient involving factorials can be costly because of the repeated evaluation. It would be beneficial if we could generate the whole vector of these probabilities in one step, i.e. without the repeated calls and even better, if this could be done instantaneously also for possibly multiple values of x denoted by a vector \mathbf{x} such that $\mathbb{Q}(t, \mathbf{k}|\mathbf{x})$ is a $(\dim(\mathbf{x}) \times \dim(\mathbf{k}))$ matrix. To this end, we propose a Fast Fourier Transform (FFT) approach (for the original FFT algorithm, see (Cooley and Tukey, 1965)) which indeed provides a significant speedup over the binomial method. The idea is to set up a probability generating function (PGF) of any credit to default. This is $p(t|x) = (1 - Q(t|x))z^0 + Q(t|x)z^1$. The PGF of the whole portfolio is $p(t|x)^n$ which is a polynomial multiplication and can be written as a convolution of $(n + 1)$ -dimensional vectors $\mathbf{p}(t|x) = [1 - Q(t|x), Q(t|x), 0, \dots, 0]$ such that $\mathbf{p}(t|x) * \dots * \mathbf{p}(t|x)$ gives the polynomial coefficients of the PGF of the number of defaults in the portfolio. The convolution, however, can be computed simply by taking FFT of the individual vectors, computing the Hamadard¹²⁶ product (\circ) of these vectors and FFT-invert the product back. Thus, the vector $\mathbb{Q}(t, \mathbf{k}|x)$ of default probabilities of \mathbf{k} names in the portfolio in terms of the FFT (and its inverse) equals

$$\mathbb{Q}(t, \mathbf{k}|x) = \mathbf{p}(t|x) * \dots * \mathbf{p}(t|x) = \text{IFFT}[\text{FFT}(\mathbf{p}(t|x)) \circ \dots \circ \text{FFT}(\mathbf{p}(t|x))], \quad (\text{IV.1.2})$$

where IFFT is the inverse operation to FFT. Although this already brings a certain speedup over the standard binomial method, as outlined above the best performance is achieved when the FFT operation is executed not only over all values of k but also over all values of x at the same time. This is in the notation analogous to the expression (IV.1.2) with the only difference that the scalar x is replaced by a column vector \mathbf{x} , so that

$$\mathbf{p}(t|\mathbf{x}) = \begin{bmatrix} 1 - Q(t|x_1) & Q(t|x_1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 - Q(t|x_{\dim(\mathbf{x})}) & Q(t|x_{\dim(\mathbf{x})}) & 0 & \cdots & 0 \end{bmatrix},$$

becomes a $(\dim(\mathbf{x}) \times (n + 1))$ matrix. It is then essential to parametrize the FFT and the IFFT such that it conducts the transformation on every row of the matrix passed in as an argument. A highly optimized MATLAB code snippet of the FFT approach is attached in the Box IV.1.5.

¹²⁶this is the original name for a common operation that is in practice referred to as 'element-wise' product. In MATLAB this operation is denoted by the dot-star operator `.*`

Box IV.1.1: MATLAB FFT algorithm for $Q(t, \mathbf{k} | \mathbf{x})$

```
% Q is a vector Q(t|x) where x is a vector
p_mat = zeros(numel(x),n+1); p_mat(:,[1 2]) = [1-Q(:) Q(:)];
% compute the matrix Q(t,k|x) by FFT a IFFT
Q_mat = ifft(fft(p_mat, [], 2).^n, [], 2);
```

We conclude that in our numerical experiment the FFT approach was about $600\times$ faster than the binomial method, when the dimension of \mathbf{x} was 500 and there were $n = 125$ credits.

IV.1.6 $Q(t|x)$ and the factor function

In the previous section it was assumed that the conditional probability of any credit to default $Q(t|x)$ is known and so we provided no further insight into its construction. This is the topic of this section. The latent one-factor model which has become a standard in the market is based on a factor function and we just quickly review the basic logic. The main idea is to assume that the log-assets of i th credit are represented by a random variable Z_i . If the value of this random variable is 'too low' then the credit defaults.¹²⁷ The 'too low' value is determined by the unconditional default probability of the credit to default. Assume $q_i(t)$ is the unconditional probability of i th credit to default until t , then

$$\tau_i^* \leq t \Leftrightarrow F_{Z_i}(Z_i) \leq q_i(t),$$

which means that i th credit is in default at time t if the CDF F_{Z_i} of the random variable Z_i evaluated at Z_i is lower or equal to the unconditional probability of default $q_i(t)$. In other words, the default is observed once the quantile of the log asset value falls below the quantile of the default time random variable. It therefore remains to define the random variable Z_i to fully describe the system. In the one-factor setup, the typical definition of Z_i is

$$Z_i = \sqrt{\rho}m_X X + \sqrt{1 - \rho}m_{Y_i} Y_i, \quad (\text{IV.1.3})$$

where X is referred to as the *common* factor, and Y_i is the *idiosyncratic* factor while both X and Y_i are mutually independent. The non-random scalars m_X, m_{Y_i} are chosen such that $\mathbb{V}[m_X X] = \mathbb{V}[m_{Y_i} Y_i] = 1$ and $\mathbb{E}[m_X X] = \mathbb{E}[m_{Y_i} Y_i] = 0$, i.e. both random variables are standardized.¹²⁸ We can substitute into Z_i the composition of the variables X, Y_i and observe when the default occurs with respect to the idiosyncratic random variable Y_i . This is expressed as the equivalence

$$\tau_i^* \leq t \Leftrightarrow Y_i \leq \frac{F_{Z_i}^{-1}(q_i(t)) - \sqrt{\rho}m_X X}{\sqrt{1 - \rho}m_{Y_i}}.$$

To obtain conditionally-independent probabilities of default, the common factor X is fixed to some value $X = x$ and then the probability of default Q_i conditionally on $X = x$, equals

$$Q_i(t|x) = F_{Y_i} \left(\frac{F_{Z_i}^{-1}(q_i(t)) - \sqrt{\rho}m_X x}{\sqrt{1 - \rho}m_{Y_i}} \right).$$

¹²⁷see the connection with the firm-value models (Part II) at this point.

¹²⁸although the standardization is not necessary in order to allow for the variables to be correlated, it is a common practice to standardize the variables.

If all the names in the basket are homogeneous, that is, have the same unconditional default probability $q_i = q$ and their idiosyncratic factors Y_i are from the same distribution, then $m_{Y_i} = m_Y$ and $Q_i \equiv Q$ which means the conditional default probability is the same across all names. This is the central idea which underpins the pricing of CDOs. The unconditional probability of default $q(t)$ is commonly assumed to be equal $q(t) = \Pr[\tau^* \leq t]$, where τ^* has the exponential distribution with intensity $\gamma = \frac{\text{index quote}}{1-\delta}$,¹²⁹ ¹³⁰ where 'index quote' is the index of the reference CDO portfolio (e.g. iTraxx or CDX.NA), and δ is the pre-specified recovery. Hence $q(t) = 1 - \exp(-\gamma t)$.

In what follows, we will choose X, Y_i to be: (i) normally distributed, which is the standard Gaussian copula model, (ii) NIG distributed which is called the 'NIG' copula and finally (iii) student t_ν -distributed which implies the double t copula model. We will also present a stochastic correlation model which can be combined with any of the choices (i)–(iii) above, yet it allows for several possible states of the copula correlation ρ . Further in the text, we also present models based on Archimedean copulas which do not have this factor representation and generate the conditional probabilities of default in a different way.

IV.1.7 The Normal Inverse Gaussian (NIG) copula

The NIG copula is even intuitively a suitable approach to CDO modeling since it is based on the 'normal' distribution, yet it is rather its generalization. The NIG distribution belongs to a family of generalized hyperbolic distributions (Barndorff-Nielsen, 1997) and as shown for example by (Kalemanova et al., 2005) or more recently by (Schlosser, 2011) and (Choroś-Tomczyk et al., 2014) it could indeed be potentially an interesting choice for the CDO modeling. NIG copula is conceptually quite similar to the Gaussian or double t copula (which is introduced later) in that it is constructed directly through NIG-distributed variables because the copula itself does not allow for closed-form representation which is available for the Archimedean copulas discussed later in the Section IV.1.10. Hence, in the case of the NIG copula, we are working with the factor function (IV.1.3) instead of an explicit formula for copula. Naturally, at the outset we need to discuss the distributional properties of the NIG distribution.

We start with the definition of a NIG-distributed random variable.

Definition IV.1.1 (NIG distribution)

Let $Y \sim \text{IG}(\delta\gamma, \gamma^2)$ be Inverse Gaussian-distributed (IG) random variable where $\gamma = \sqrt{\alpha^2 - \beta^2}$. Then

$$X \triangleq (X|Y = y) \sim N(\mu + \beta y, y) \text{ with } 0 \leq |\beta| \leq \alpha, \mu \in \mathbb{R}, \delta > 0,$$

has the NIG($\alpha, \beta, \mu, \delta$) distribution. The PDF f_X and the CF φ_X , respectively, have the

¹²⁹this relationship is often known under the term 'credit triangle'. Although being simplification of a more complicated relation, it works surprisingly well for many problems. See (Kane, 2008, Section 3.10) for a more detailed discussion.

¹³⁰although this expression is very simplifying, see the connection with the reduced-form models presented in the Part III.

forms

$$f_X(x) = \frac{\delta\alpha \exp(\delta\gamma + \beta(x - \mu))}{\pi\sqrt{\delta^2 + (x - \mu)^2}} K_1\left(\alpha\sqrt{\delta^2 + (x - \mu)^2}\right)$$

$$\varphi_X(u) = \exp\left(\delta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + iu)^2}\right) + \mu iu\right).$$

Here, $K_1(x)$ is the modified Bessel function of the third kind $K_I(x)$ with index $I = 1$.¹³¹ It is thus easy to see from the Definition IV.1.1 that the NIG random variable is actually a Normally distributed random variable with mean and variance subject to the IG distribution. Since both the Normal and the IG distribution are analytically well tractable distributions, it can be expected that the NIG distribution possess at least some degree of the analytical tractability too. And this is indeed the case as both the PDF f_X and the CF φ_X have closed-forms, although the CDF is not known in the closed-form. The parameter α controls the tail heaviness, $\beta = 0$ ensures a symmetricity of the distribution otherwise it is non-symmetrical, μ is a location parameter and δ is a scale parameter. From the CF φ_X it is not too hard to obtain the mean and variance of NIG-distributed variable X . We can conveniently use the Moment-generating function (MGF) – CF identity which is $M_X(u) = \varphi_X(-iu)$. This means the MGF M_X has the form

$$M_X(u) = \exp\left(\delta\left(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + u)^2}\right) + \mu u\right).$$

Given the k th raw moment is $\mathbb{E}[X^k] = \frac{d^k}{du^k} M_X(0)$ it is straightforward to obtain the expected value and the variance¹³² of $X \sim \text{NIG}$

$$\mathbb{E}[X] = \mu + \frac{\beta\delta}{\gamma}, \mathbb{V}[X] = \frac{\alpha^2\delta}{\gamma^3}. \quad (\text{IV.1.4})$$

Because obviously the factor function (IV.1.3) involves the summation of the random variables and also scaling we need to check these properties of the NIG distribution. We outline both these properties in a Theorem.

Theorem IV.1.1 (Scaling and summation property of the NIG distribution)

Let $X \sim \text{NIG}(\alpha, \beta, \mu_X, \delta_X)$ and let $c > 0$. Then $cX \sim \text{NIG}\left(\frac{\alpha}{c}, \frac{\beta}{c}, c\mu_X, c\delta_X\right)$.

Let furthermore $Y \sim \text{NIG}(\alpha, \beta, \mu_Y, \delta_Y)$ be another random NIG variable independent of X . Then the sum $(X + Y)$ is also NIG-distributed random variable such that $(X + Y) \sim \text{NIG}(\alpha, \beta, \mu_X + \mu_Y, \delta_X + \delta_Y)$.

¹³¹the name of the Bessel function differs in various texts and software. For example in (Barndorff-Nielsen, 1997) and (Schlosser, 2011) it is the modified Bessel of the *third* kind, while in MATLAB it is known under the modified Bessel function of the *second* kind `besselk(1, x)`. To prevent any confusion, by writing $K_1(x)$ we mean the Bessel function defined as

$$K_1(x) = \frac{1}{2} \int_0^\infty \exp\left(-\frac{1}{2}x\left(s + \frac{1}{s}\right)\right) ds,$$

regardless of its actual name in the literature or software.

¹³²the precise steps for the variance follow the standard statistical machinery and here it is presented for the completeness. First compute the first and the second raw moments $\mathbb{E}[X] = \frac{d}{du} M_X(0)$, $\mathbb{E}[X^2] = \frac{d^2}{du^2} M_X(0)$ and then apply $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Proof. Both properties can be easily verified by the use of the CF. For the proof of the scaling property it is first needed to apply the CF identity $\varphi_{cX}(u) = \varphi_X(cu)$. This yields an expression of a similar form to that of φ_X , yet we can see that $\alpha \rightarrow \frac{\alpha}{c}, \beta \rightarrow \frac{\beta}{c}, \mu_X \rightarrow c\mu_X, \delta_X \rightarrow c\delta_X$, where the arrow ' \rightarrow ' has the meaning of 'is replaced by'. Hence, we can conclude that $cX = \tilde{X}$ where \tilde{X} is defined as $\tilde{X} \sim \text{NIG}\left(\frac{\alpha}{c}, \frac{\beta}{c}, c\mu_X, c\delta_X\right)$. To prove the summation property it is enough to apply $\varphi_{X+Y}(u) = \varphi_X(u)\varphi_Y(u)$ and check that $\varphi_{X+Y}(u)$ has the same form as CF of any NIG random variable with parameters $\alpha, \beta, \mu_X + \mu_Y, \delta_X + \delta_Y$. \square

With the useful properties listed in the Theorem IV.1.1 we can finally take a closer look at the factor function (IV.1.3) where both the scaling and the summation properties are needed. In what follows we assume that both the standardizing scalars m_X, m_Y are set to 1 as we can control the variance of X, Y_i by the choice of the parametrization. This means we are working with the factor function

$$Z_i = \sqrt{\rho}X + \sqrt{1 - \rho}Y_i.$$

As we are about to proceed from general to specific cases, at the outset we assume $X \sim \text{NIG}(\alpha_X, \beta_X, \mu_X, \delta_X), Y_i \sim \text{NIG}(\alpha_Y, \beta_Y, \mu_Y, \delta_Y)$, that is X and Y_i both are NIG-distributed yet the distributional parameters of X and Y_i are in general different and the idiosyncratic factors $Y_i, i = 1, \dots, n$ have all the same NIG distribution. There are three main objectives to be completed. Firstly, we want to identify the distribution of Z_i , secondly we need to ensure that variables X, Y_i are standardized,¹³³ and as a third, we would like to reduce the amount of parameters involved.

The *first* task involves application of both properties of the Theorem IV.1.1. After scaling X by $\sqrt{\rho}$ and Y_i by $\sqrt{1 - \rho}$ we are trapped by the limiting fact that two NIG random variables need to share the first two parameters in order that their sum is also NIG-distributed. This can be circumvented by imposing a restriction on the parameter choice and after reflecting this restriction, we set $\alpha \equiv \alpha_X, \alpha_Y = \alpha\sqrt{\frac{1}{\rho} - 1}, \beta \equiv \beta_X, \beta_Y = \beta\sqrt{\frac{1}{\rho} - 1}$. To meet the first objective, we thus need to have

$$\begin{aligned} X &\sim \text{NIG}(\alpha, \beta, \mu_X, \delta_X) \\ Y_i &\sim \text{NIG}\left(\alpha\sqrt{\frac{1}{\rho} - 1}, \beta\sqrt{\frac{1}{\rho} - 1}, \mu_Y, \delta_Y\right) \\ Z_i &\sim \text{NIG}\left(\frac{\alpha}{\sqrt{\rho}}, \frac{\beta}{\sqrt{\rho}}, \sqrt{\rho}\mu_X + \sqrt{1 - \rho}\mu_Y, \sqrt{\rho}\delta_X + \sqrt{1 - \rho}\delta_Y\right). \end{aligned}$$

To satisfy the *second* requirement which states to have the standardized variables X, Y_i , we in fact need to ensure $\mathbb{E}[X] = \mathbb{E}[Y_i] = 0, \mathbb{V}[X] = \mathbb{V}[Y_i] = 1$ by again imposing another restriction on the parametrization of X, Y_i .¹³⁴ This restriction on the parameters is based on a system of equations defining zero mean and unit variance for both X, Y_i (see (IV.1.4) for the mean and the variance of the NIG random variable). Solving the system

¹³³this was already outlined above that we want all the factor variables to be standardized.

¹³⁴notice that $\mathbb{E}[Z_i] = 0, \mathbb{V}[Z_i] = 1$ automatically when X, Y_i are standardized.

of equations results into

$$\begin{aligned}
X &\sim \text{NIG} \left(\alpha, \beta, -\frac{\beta\gamma^2}{\alpha^2}, \frac{\gamma^3}{\alpha^2} \right) \\
Y_i &\sim \text{NIG} \left(\alpha\sqrt{\frac{1}{\rho}-1}, \beta\sqrt{\frac{1}{\rho}-1}, -\sqrt{\frac{1}{\rho}-1}\frac{\beta\gamma^2}{\alpha^2}, \sqrt{\frac{1}{\rho}-1}\frac{\gamma^3}{\alpha^2} \right) \\
Z_i &\sim \text{NIG} \left(\frac{\alpha}{\sqrt{\rho}}, \frac{\beta}{\sqrt{\rho}}, -\frac{1}{\sqrt{\rho}}\frac{\beta\gamma^2}{\alpha^2}, \frac{1}{\sqrt{\rho}}\frac{\gamma^3}{\alpha^2} \right), \gamma = \sqrt{\alpha^2 - \beta^2}.
\end{aligned}
\tag{IV.1.5}$$

The *third* objective was the reduction of the number of variables in the system. This has partially already been achieved as after tackling the second objective, the number of parameters in the system has reduced from the original 9 to only 3.¹³⁵ It is, however, easy to see that a further reduction can be obtained if we force the variables X, Y_i, Z_i to be generated by a *symmetrical* NIG distribution. As the (a)symmetry of the NIG distribution is exclusively controlled by the second parameter, the only way to ensure this symmetry in (IV.1.5), is to set $\beta = 0$. After setting $\beta = 0$ in (IV.1.5), we finally have

$$\begin{aligned}
X &\sim \text{NIG}(\alpha, 0, 0, \alpha) \\
Y_i &\sim \text{NIG} \left(\alpha\sqrt{\frac{1}{\rho}-1}, 0, 0, \alpha\sqrt{\frac{1}{\rho}-1} \right) \\
Z_i &\sim \text{NIG} \left(\frac{\alpha}{\sqrt{\rho}}, 0, 0, \frac{\alpha}{\sqrt{\rho}} \right),
\end{aligned}
\tag{IV.1.6}$$

which is the most simplified, two-parameter (α, ρ) -setup for the NIG copula model.

The philosophical question is whether to use the 'richer' setup (IV.1.5) or the simpler one in (IV.1.6) for modeling. The advantages and disadvantages are here clearly given. The former setup is endowed with a greater flexibility as three parameters (α, β, ρ) govern the distributional properties of the system, one of which (β) also controls the (a)symmetry of the distributions. The argument against is that for the number of quotes provided, which are usually 4 up 6 (cf. Table 15) this might lead to an overfitting. The latter specification given by (IV.1.6) assumes $\beta = 0$ and so there are only two parameters to be set (α, ρ) . This makes the setup simpler, prevents overfitting yet it might not be flexible enough to fit the market quotes properly. We will investigate both these setups in the Section IV.1.11 to check which one is more appropriate for the practical use. Finally, notice there are also some technical issues with the implementation of the NIG model as neither the CDF of NIG distribution nor inverse CDF is known in an analytical form.

IV.1.8 The double t copula

The double t copula is a model based on the t distribution which for finite degrees of freedom is known to have heavier tails than the normal distribution as a benchmark. In

¹³⁵at the outset there were 4 parameters for X , another 4 parameters for Y_i and a correlation ρ . After the second step there are only three parameters α, β, ρ in the system. Notice γ is not an extra parameter as $\gamma = \sqrt{\alpha^2 - \beta^2}$.

turn, we would expect the t distribution as a 'generalization' of the normal distribution to produce more consistent prices of credit products than the normal distribution. The double t copula model was pioneered by (Hull and White, 2004), who found out that the double t specification is suitable for CDO pricing yet the analysis was done on a limited data sample which coincidentally could have been appropriate for the double t model. Later, the double t model was discussed also in a greater detail in (Vrins, 2009) for structured products. (Kalemanova et al., 2005) also applies the double t approach and confirms that it has some desired properties for the CDO modeling but still she works with the data that are pre-crisis that are more or less similar to the study of (Hull and White, 2004). All the mentioned works find the double t model suitable for the CDO pricing, yet they do not provide much insight into the technical aspects of the double t model and apply the simulation-based approach which is easy to be set up. The simulation, however, is a totally unsuitable technique for the calibration, yet this is a critical component of a complex CDO pricing. In this section, we provide an exhaustive treatment of the double t model with a particular focus on the implementation as the technical complications are arguably the most discouraging part of the model. We also find that for some special cases the double t model is almost analytically tractable.

Independently on the other authors, to build-up the model, we proceed as follows. First, we define the generalized case, where the factors do not originate from a t distribution with the *same* degrees of freedom. Then, we focus on the more common specification where both the systematic and the idiosyncratic factors (X and Y_i , respectively) originate from the t distribution with the same degrees of freedom parameter ν .

To start with the **general** case, assume that $X \sim t_{\nu_X}$ (systematic factor) and $Y_i \sim t_{\nu_Y}$, $i = 1, \dots, n$ (idiosyncratic factors) and $\nu_X, \nu_Y > 2$.¹³⁶ Having defined the distribution of the factors, the factor function (IV.1.3) has the form

$$Z_i = \sqrt{\rho} \sqrt{\frac{\nu_X - 2}{\nu_X}} X + \sqrt{1 - \rho} \sqrt{\frac{\nu_Y - 2}{\nu_Y}} Y_i, \quad (\text{IV.1.7})$$

where the terms $m_X = \sqrt{\frac{\nu_X - 2}{\nu_X}}$, $m_{Y_i} = \sqrt{\frac{\nu_Y - 2}{\nu_Y}}$ are standardizing factors so that $\mathbb{V} \left[\sqrt{\frac{\nu_X - 2}{\nu_X}} X \right] = \mathbb{V} \left[\sqrt{\frac{\nu_Y - 2}{\nu_Y}} Y_i \right] = 1$ and in turn also $\mathbb{V}[Z_i] = 1$.¹³⁷ From the known properties of the t distribution the main complication arises with the random variable Z_i as this is no longer a t -distributed variable and does not even have any other known distribution. We thus need some other, alternative way of describing the distribution of Z_i in (IV.1.7). We present this main idea in a Theorem.

Theorem IV.1.2 (The distribution of Z_i)

The random variable Z_i from (IV.1.7) has a distribution whose characteristic function is

$$\varphi_{Z_i}(u) = \eta |u|^{\frac{1}{2}(\nu_X + \nu_Y)} K_{\frac{\nu_X}{2}} \left(\sqrt{(\nu_X - 2)\rho} |u| \right) K_{\frac{\nu_Y}{2}} \left(\sqrt{(\nu_Y - 2)(1 - \rho)} |u| \right),$$

with

$$\eta = \frac{2^{2 - \frac{1}{2}(\nu_X + \nu_Y)} ((\nu_X - 2)\rho)^{\frac{\nu_X}{4}} ((\nu_Y - 2)(1 - \rho))^{\frac{\nu_Y}{4}}}{\Gamma\left(\frac{\nu_X}{2}\right) \Gamma\left(\frac{\nu_Y}{2}\right)}.$$

¹³⁶the inequalities $\nu_X, \nu_Y > 2$ ensure that X, Y_i have a finite variance.

¹³⁷it is required that only the variance is set to 1 because the mean of the t -distributed variable is 0 by definition.

The function $K_I(x)$ is the same Bessel function as in the Definition IV.1.1.

Proof. It was shown by (Hurst, 1997) that the t distribution has a closed-form CF in terms of the Bessel function $K_I(x)$. This allows to find the CFs φ_X, φ_{Y_i} of the t_{ν_X}, t_{ν_Y} -distributed random variables X, Y_i , respectively in (IV.1.7). Because furthermore $\varphi_{\sqrt{\rho \frac{\nu_X - 2}{\nu_X}} X}(u) = \varphi_X\left(u \sqrt{\rho \frac{\nu_X - 2}{\nu_X}}\right), \varphi_{\sqrt{(1-\rho) \frac{\nu_Y - 2}{\nu_Y}} Y_i}(u) = \varphi_{Y_i}\left(u \sqrt{(1-\rho) \frac{\nu_Y - 2}{\nu_Y}}\right)$, we can compute

$$\varphi_{Z_i}(u) = \varphi_{\sqrt{\rho \frac{\nu_X - 2}{\nu_X}} X + \sqrt{(1-\rho) \frac{\nu_Y - 2}{\nu_Y}} Y_i}(u) = \varphi_X\left(u \sqrt{\rho \frac{\nu_X - 2}{\nu_X}}\right) \varphi_{Y_i}\left(u \sqrt{(1-\rho) \frac{\nu_Y - 2}{\nu_Y}}\right),$$

leading to the expression for $\varphi_{Z_i}(u)$ in the Theorem IV.1.2 which concludes the proof. \square

Once the characteristic function φ_{Z_i} is established it allows us to compute all the distributional quantities required for the pricing. In fact, we need the inverse CDF $F_{Z_i}^{-1}$ as this function is used in the computation of $Q(t|x)$ (see the Section IV.1.6) needed for the CDO pricing. In order to evaluate the quantile function $F_{Z_i}^{-1}(p)$ at $p \in (0, 1)$ we will use a root-searching algorithm which calculates the p -quantile value q such that $q = \{z \in \mathbb{R} : F_{Z_i}(z) - p = 0\}$. A Newton-Raphson can effectively be used to find q . In turn the problem of computing $F_{Z_i}^{-1}$ has switched to a problem of calculating F_{Z_i} several times. Although the CDF F_{Z_i} is unknown too, there is already an efficient way of evaluating it numerically given the CF φ_{Z_i} . This *CF to CDF* transformation is known under the Gil-Pelaez transform¹³⁸ (see (Gil-Pelaez, 1951) or for alternative, yet numerically equivalent form (Shephard, 1991)) and it reads

$$\begin{aligned} F_{Z_i}(z) &= \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \operatorname{Re} \left\{ \frac{\exp(-iuz) \varphi_{Z_i}(u)}{iu} \right\} du \\ &= \frac{1}{2} - \frac{\eta}{\pi} \\ &\quad \times \int_0^{\infty} \operatorname{Re} \left\{ \frac{e^{-iuz} |u|^{\frac{1}{2}(\nu_X + \nu_Y)} K_{\frac{\nu_X}{2}}\left(\sqrt{(\nu_X - 2)\rho}|u|\right) K_{\frac{\nu_Y}{2}}\left(\sqrt{(\nu_Y - 2)(1-\rho)}|u|\right)}{iu} \right\} du, \end{aligned}$$

where we already substituted the CF φ_{Z_i} . Notice that when calculating the quantile value q the numerical root-finding algorithms usually require (or perform better) that an initial guess q^* of q is provided. We recommend to use q^* from a t -distribution with $\nu = (\sqrt{\rho\nu_X} + \sqrt{1-\rho\nu_Y})$ which heuristically expresses the idea that the distribution of Z_i is likely to be 'close' to the t distribution with the number of degrees of freedom given by a linear combination of the number of degrees of freedom of two factors X, Y_i involved. Although this is not perfect, as an initial guess it performs very well. Computationally, it is much less costly to compute the inverse CDF of the t distribution and then to have only a few iterations for finding q than to skip the initial guess and iterate many times to find q . The more information is provided at the outset, the faster the calculations will be.

¹³⁸notice the Gil-Pelaez transform was already in place when we were dealing with the Heston model in the Section II.1.4.4. However, in the Heston model we were not really interested in how the main Heston's result was found so the main ideas behind the Gil-Pelaez transform was rather in the background.

Now that the generalized double t copula model has been established, we can introduce a **simplified** version. The simplification is intuitive. Instead of allowing $\nu_X \neq \nu_Y$, we set $\nu_X = \nu_Y \equiv \nu > 2$ which means both factors X and Y_i are assumed to have the same t distribution with $\nu > 2$ degrees of freedom. Repeating the steps above gives the simpler CF

$$\varphi_{Z_i}(u) = \frac{2^{2-\nu} \left((\nu-2)^2 (1-\rho) \rho \right)^{\frac{\nu}{4}} |u|^\nu}{\Gamma\left(\frac{\nu}{2}\right)^2} K_{\frac{\nu}{2}} \left(\sqrt{(1-\rho)(\nu-2)} |u| \right) K_{\frac{\nu}{2}} \left(\sqrt{\rho(\nu-2)} |u| \right).$$

We also found that there exists a **special** case when the Bessel function has an expression involving only elementary functions. This happens when ν is an odd number. We found the following expression for $K_{\frac{\nu}{2}}(x)$ given ν is *odd*:

$$K_{\frac{\nu}{2}}(x) = \frac{e^{-x} \sqrt{\frac{\pi}{2}}}{\sqrt{x}} \left(1 + \frac{(\nu-2)!!}{x^{\frac{\nu-1}{2}}} + \frac{(\nu-2)!!}{x^{\frac{\nu-2-1}{2}}} + \frac{(\nu-4)!!}{x^{\frac{\nu-4-1}{2}}} + \frac{(\nu-6)!!}{x^{\frac{\nu-6-1}{2}}} + \dots + \frac{\dots}{x} \right), \nu = 3, 5, \dots, \quad (\text{IV.1.8})$$

where '!!' is the double factorial of an odd number.¹³⁹ In this expression it is critical to follow the rule that the series involves terms ending by $\frac{\dots}{x}$, that is the last term in the series always has x in the denominator. The benefit of this special case representation is that when used in the CF φ_{Z_i} the CF has a nice form from which it is even possible to obtain closed-forms expression for both the PDF f_{Z_i} and the CDF F_{Z_i} by applying the Fourier Inversion theorem. This is not new, as we have already applied the Fourier Inversion theorem in the Proposition IV.1.1. In this case, we can proceed similarly, just with φ_{Z_i} instead of φ .

In the case of the double t copula model, we are facing the same dilemma as in the NIG copula model, that is the choice of the model version. Here, we can choose from a three-parameter (ν_X, ν_Y, ρ) variant or a simpler, two-parameter case (ν, ρ) . We test both cases to conclude which implementation is more meaningful for the practical pricing purposes.

IV.1.9 The stochastic correlation model

So far, it was assumed that the correlation parameter ρ is a fixed number. This simplifying assumptions is, however, not a true reflection of the reality as the recent research shows the correlations of market factors tighten during the stress period and tend to lower during the market standstill. For example, (Longin and Solnik, 2001) use the hypothesis testing approach and conclude that correlation is insignificant in bullish markets, yet increases when the the markets are bearish. (Loretan and English, 2000) show that in the case of the market downturn the correlation tend to be higher relative to the 'what the theory suggests' and provide implications for the regulation. A comprehensive study (Sandoval and Franca, 2012) investigates the linkages between the volatility in the markets and correlations using the principal component analysis applied to various crisis periods (1987, 1998, 2001 and 2008). The authors conclude that there is a significant relationship between the market volatility and the correlation. Also, (Christoffersen et al., 2013) argue that dependencies among credits are dynamical and aim to dynamize

¹³⁹for example $7!! = 7 \times 5 \times 3 \times 1 = 105$.

the copula. All the listed studies agree on the fact that the correlation changes with periods and hence it is too restrictive to assume it is a constant. It would be therefore desirable that a comprehensive model can account for *regime-switching correlation*. In this we follow (Gregory et al., 2007) who introduce a simple yet practical approach to regime-switching correlation in a factor function.

The main idea is to modify the factor function (IV.1.3) by adding two mutually independent Bernoulli-distributed random variables B_a, B_b so that conditionally on $B_a, B_b \in \{0, 1\}$ the factor function either retains its 'original' form (IV.1.3), or it is given by $Z_i = X$ or finally is stated as $Z_i = Y_i$. The last two artificially added states are equivalent to the cases when $\rho = 1$ and $\rho = 0$, respectively. We can thus conclude that adding B_a, B_b the factor function behaves as if the copula correlation were stochastic and either attained the value $\rho, 1$ or 0 . In order to make this work, the factor function has to be specified as

$$Z_i = (1 - B_a)(1 - B_b) \left(\sqrt{\rho} m_X X + \sqrt{1 - \rho} m_{Y_i} Y_i \right) + (1 - B_a) B_b m_{Y_i} Y_i + B_a m_X X.$$

Based on this formulation, three non-overlapping scenarios for Z_i can occur with respect to the states of B_a, B_b :

$$Z_i = \begin{cases} m_X X & \text{if } B_a = 1 \\ m_{Y_i} Y_i & \text{if } B_a = 0, B_b = 1 \\ \sqrt{\rho} m_X X + \sqrt{1 - \rho} m_{Y_i} Y_i & \text{if } B_a = 0, B_b = 0 \end{cases}.$$

Since the main objective is to compute the $X = x$ -conditional probability of default, we are interested in how the probability $Q(t|x)$ is defined given this extended specification where the correlation in Z_i can take on several values. Denote $p_a = \Pr[B_a = 1], p_b = \Pr[B_b = 1]$. Then, the function $Q(t|x)$ takes the following form

$$Q(t|x) = \begin{cases} \mathbf{1}_{\{m_X x \leq F_{Z_i}^{-1}(q_i(t))\}} & \text{with probability } p_a \\ F_{Y_i} \left(\frac{F_{Z_i}^{-1}(q_i(t))}{m_{Y_i}} \right) & \text{with probability } (1 - p_a)p_b \\ F_{Y_i} \left(\frac{F_{Z_i}^{-1}(q_i(t)) - \sqrt{\rho} m_X x}{\sqrt{1 - \rho} m_{Y_i}} \right) & \text{with probability } (1 - p_a)(1 - p_b) \end{cases}.$$

Here, we used simply the quantile-to-quantile specification of default

$\tau_i^* \leq t \Leftrightarrow Z_i \leq F_{Z_i}^{-1}(q_i(t))$. It can be seen that in the first scenario, surprisingly, the default is no longer a random variable because $X = x$ is a fixed factor, and there is no space for Y_i which means the randomness has been ruled out and the default event is a completely deterministic event represented by the indicator $\mathbf{1}_{\{\cdot\}}$. Collecting all the scenarios into just one expression then finally gives

$$\begin{aligned} Q(t|x) &= p_a \mathbf{1}_{\{m_X x \leq F_{Z_i}^{-1}(q_i(t))\}} + (1 - p_a)p_b F_{Y_i} \left(\frac{F_{Z_i}^{-1}(q_i(t))}{m_{Y_i}} \right) \\ &+ (1 - p_a)(1 - p_b) F_{Y_i} \left(\frac{F_{Z_i}^{-1}(q_i(t)) - \sqrt{\rho} m_X x}{\sqrt{1 - \rho} m_{Y_i}} \right). \end{aligned}$$

It is thus easy to see that the $X = x$ -conditional probability $Q(t|x)$ is a p_a, p_b -weighted combination of $X = x$ -conditional probabilities of default given for a particular scenario of the factor function Z_i .

The beauty of this special regime-switching, or stochastic correlation specification is in that it can be seamlessly combined with any factor function such as the one for Gaussian, NIG (Section IV.1.7), double t (Section IV.1.8) copula and hence this extension is quite general. One, however, needs to keep in mind that there are two extra parameters p_a, p_b and so the total number of the parameters in the system is equal to the number of parameters from the factor function + 2. In the models we analyzed, this may lead up to 5 parameters in the system which could make the model overparametrized.¹⁴⁰ This stochastic correlation extension is therefore rather used in combination with the Gaussian copula so that the parameters to be estimated are (ρ, p_a, p_b) .

IV.1.10 $Q(t|x)$ and the Archimedean copulas

In this section, we outline the approach of CDO pricing via one-parameter (θ) Archimedean copulas. Under this approach, the construction of conditionally-independent probabilities of default $Q_i(t|x)$ differs from that of discussed in the previous Section IV.1.6 in that there is no 'correlation' present in the Archimedean copulas and hence the construction lacks the factor function Z_i , yet as we show the conditional independence can still be achieved. Notice that in the case of the Normal, NIG and double t copulas we were modeling the log-asset variable Z_i by coupling together the variables X, Y_i following the same distribution (with possibly different parameters). This was necessary as there was no analytical 'copula' function C that would produce the same dependency structure between some Z_i, Z_j we were able to obtain through the factor function. The absence of the factor function Z_i in the case of the Archimedean copulas means that there is no log-asset variable but the probabilities of default can be modeled directly through the Archimedean copula C which has an analytical form.

Assume there exists a *jointly* exponentially distributed vector of random default times $(\tau_1^*, \dots, \tau_n^*)$ with some CDF H . According to the Sklar's theorem (see e.g. (Tapiero, 2013, p. 132)), we observe

$$\begin{aligned} H(t_1, \dots, t_n) &= \Pr[\tau_1^* \leq t_1, \dots, \tau_n^* \leq t_n] \\ &= \Pr[q_1(\tau_1^*) \leq q_1(t_1), \dots, q_n(\tau_n^*) \leq q_n(t_n)] \\ &= \Pr[U_1 \leq q_1(t_1), \dots, U_n \leq q_n(t_n)] \\ &= C(q_1(t), \dots, q_n(t)), \end{aligned}$$

where C is the copula function. Hence, we observe default of the name i by time t , iff $U_i \leq q_i(t)$ where the uniforms $U_i, i = 1, \dots, n$ come from the *copula distribution*. This is a general result and was implicitly applied also in the Section IV.1.6.¹⁴¹ The main question now is how to generate the vector (U_1, \dots, U_n) whose dependency structure is given by the

¹⁴⁰recall that the usual calibration set only contains 4-6 quotes. An example is shown in the Table 15.

¹⁴¹notice $U_i = F_{Z_i}(Z_i) = F_{Z_i}(\sqrt{\rho}m_X X + \sqrt{1-\rho}m_{Y_i} Y_i)$.

copula C . Moreover, we also want to compute the default probability $\Pr[U_i \leq q_i(t)]$ given the dependency structure between the set of $U_i, i = 1, \dots, n$ by the copula C . Both these issues are tackled by the Marshall-Olkin method of generating random uniform variables U_i having the Archimedean copula C . We refer the reader to (McNeil et al., 2005) for a general discussion about copulas.

The Archimedean copulas (see e.g. (Ruppert, 2011)) are a class of copulas which are defined explicitly through the copula expression

$$C(u_1, \dots, u_n) = \phi^{-1}(\phi(u_1) + \dots + \phi(u_n)),$$

where ϕ is a *generator* function satisfying certain conditions. We will assume the following two functional forms of the Archimedean copula generator ϕ :

$$\phi(u) = \begin{cases} \frac{1}{\theta} \left(\frac{1}{u^\theta} - 1 \right), \theta > 0 & \text{for the Clayton copula} \\ (-\ln u)^\theta, \theta \geq 1 & \text{for the Gumbel copula} \end{cases}. \quad (\text{IV.1.9})$$

To simulate $U_i, i = 1, \dots, n$ from the dependency structure given by the Archimedean copula C , Marshall-Olkin (Marshall and Olkin, 1988) suggest to take the following steps:

1. Generate a random variable X , from a distribution whose CDF is F such that the Laplace-Stieltjes transform $\tilde{F}(z) = \int_0^\infty \exp(-zx) dF(x)$ of F is equal to the inverse of the generator ϕ , so $\tilde{F}(z) = \phi^{-1}(z)$.
2. Draw mutually independent uniforms $Y_i \sim U[0, 1], i = 1, \dots, n$.
3. Construct the uniform random variables $U_i, i = 1, \dots, n$ having jointly the C -copula distribution as

$$U_i = \tilde{F} \left(-\frac{\ln Y_i}{X} \right). \quad (\text{IV.1.10})$$

It is thus easy to see that *conditionally on* $X = x, U_i, i = 1, \dots, n$ are independent uniformly distributed variables from the copula C . To compute U_i using this method, it remains to find $\tilde{F} = \phi^{-1}$ and the connected function F that corresponds to the generators in (IV.1.9) for the Clayton and Gumbel copulas. A straightforward yet tedious computation shows that

$$\tilde{F}(z) = \phi^{-1}(z) = \begin{cases} (1+z)^{-\frac{1}{\theta}} \Rightarrow F = F_{\Gamma(\frac{1}{\theta}, 1)} & \text{for the Clayton c.} \\ \exp\left(-z^{\frac{1}{\theta}}\right) \Rightarrow F = F_{\text{St}(\frac{1}{\theta}, 1, [\cos(\frac{\pi}{2\theta})]^\theta, 0)} & \text{for the Gumbel c.} \end{cases},$$

where $\Gamma(\frac{1}{\theta}, 1)$ is a Gamma distribution with shape parameter $\frac{1}{\theta}$, scale parameter 1 and $\text{St}(\frac{1}{\theta}, 1, [\cos(\frac{\pi}{2\theta})]^\theta, 0)$ is a Stable distribution with parameter of stability, skewness, scale and location, respectively.

As mentioned earlier in this section, our objective was to infer a formula for the probability of default $\Pr[U_i \leq q_i(t)]$ of any name i by letting U_i come from the copula C distribution. Substituting (IV.1.10) into $\Pr[U_i \leq q_i(t)]$ and fixing the common factor $X = x$, we find

$$Q_i(t|x) = \Pr[U_i \leq q_i(t) | X = x] = \Pr \left[\tilde{F} \left(-\frac{\ln Y_i}{x} \right) \leq q_i(t) \right].$$

Using the functions \tilde{F} above and expressing the probability of default with respect to the only random variable Y_i , we finally have

$$Q_i(t|x) = \begin{cases} \exp\left(x\left(1 - q_i(t)^{-\theta}\right)\right) & \text{for the Clayton copula} \\ \exp\left(-x\left(-\ln q_i(t)\right)^\theta\right) & \text{for the Gumbel copula} \end{cases}.$$

As in the CDO model all the credits are assumed to be homogeneous, we can furthermore drop the subscript i .

At the end of this section, let us discuss the technical issue related to the Stable distribution. The difficulty here is that the common factor X has to be eventually integrated out as discussed at the end of the Section IV.1.4, yet the PDF f_X of the Stable distribution is not defined. We thus have to resort to a characteristic function φ and compute the density numerically by the Fourier inversion theorem.

Proposition IV.1.1

A random variable from the Stable distribution $\text{St}(\alpha, \beta, c, \mu)$ has the PDF

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \varphi(u) du, \text{ where}$$

$$\varphi(u) = \exp\left(iu\mu - |cu|^\alpha \left(1 - i\beta \text{sgn}(u) \tan\left(\frac{\alpha\pi}{2}\right)\right)\right), \alpha \neq 1.$$

For further properties the reader is referred to (Rachev, 2004, Section 8). Although this density function might be complicated to be repeatedly evaluated for various x , a smart implementation of the integration can produce very fast results. A particularly significant speedup is achieved once the Gauss-Legendre quadrature is applied and the quadrature is adjusted so that \mathbf{x} is a vector and then $f(\mathbf{x})$ is evaluated. A highly optimized MATLAB implementation is shown in the Box IV.1.10.

Box IV.1.2: MATLAB code for fast PDF of Stable distribution

```
function fx = stable_PDF(x, u, w, alpha, beta, c, mu)
% function computes PDF fx = f(x), where x is a vector of points
% u,w = integration points and weights for Gaussian quadrature
    u = u(:); w = w(:); x = x(:);
    x_mat = repmat(x, 1, numel(w)); w_mat = repmat(w', numel(x), 1);
    CF = exp(1i*u*mu-abs(c*u).^alpha.*(1-1i*beta*sign(u).*tan(alpha*pi/2)));
    CF = repmat(CF', numel(x), 1);
    fx = w_mat.* (exp(-1i*repmat(u', numel(x), 1).*x_mat).*CF);
    fx = 1/(2*pi)*real(sum(fx, 2));
end
```

Considering that $\dim(\mathbf{x})$ is 1000 (points), and we are using 300-point quadrature, this algorithm calculates $f(\mathbf{x})$ in about 0.02s which is a 50× speed gain against a simple implementation computing $f(x)$ for every single point x of \mathbf{x} .

IV.1.11 The data, calibration procedure and assessment of the models

For the selection of the optimal model, data samples are needed. We have collected 4 data sets of iTraxx 5Y CDO single tranche quotes from different periods (2008, 2013, 2014

and 2015), where intentionally the 2008 crisis dataset is included. We calibrate the model to the market data using the least-squares procedure by fitting the market-quoted prices (spreads/upfronts¹⁴²) of the tranches. Let $V_j(\Theta)$ denote the model price of j th the CDO tranche (the total number of tranches in the data snapshot is J) given a parameter set Θ and let V_j^M be the market-quoted price of j th tranche. The task is to find the optimal set of parameters $\hat{\Theta}$ such that

$$\hat{\Theta} = \arg \min_{\Theta} [\varepsilon(\Theta)], \varepsilon(\Theta) = 10000 \times \sum_{j=1}^J (V_j(\Theta) - V_j^M)^2,$$

where we added the 10000 multiplier to magnify the small numbers. The assessment of the models is a complex issue as there are several criteria that can be assessed. The crucial metric is the goodness-of-fit since the main motivation is the pricing of the bespoke tranches. For every dataset, we want a *single* set of parameters $\hat{\Theta}$ to match the market quotes. As we have presented various approaches of which some just differ by the number of parameters in the model, we also take into account the number of parameters to achieve the fit. The less number of parameters are needed to achieve a good fit, the better *ceteris paribus* the model. Ideally, we would also like the model to produce unique prices with respect to the correlation parameter ρ . That is, the tranche spread should be a monotonic function of the correlation parameter ρ .¹⁴³

The market quotes (premiums) and the model premiums for the 4 datasets are shown in the Tables 16–19. Each row shows quotes for a given tranche, sorted from the equity tranche (0%-3%) up to the (super) senior tranche (12%-22% or 22%-100%). The first column contains prices quoted by the dealers and the other columns show prices generated by the calibrated models. The labelling of the models is as follows. 'Normal' denotes the classical Gaussian copula model, 'Normal+' is a gaussian copula with stochastic correlation presented in the Section IV.1.9. 'NIG' is the NIG copula model with only (α, ρ) being fitted. 'NIG+' is the NIG copula where (α, β, ρ) are being fitted. 'Double t ' corresponds to the double t copula, where $\nu_X = \nu_Y = \nu$, that is both factors X and Y_i have the same t distribution. 'Double $t+$ ' then assumes the double t copula model with $\nu_X \neq \nu_Y$, i.e. the factors X, Y_i are allowed to be drawn from t distributions with different degrees of freedom. 'Clayton' and 'Gumbel' denote the two Archimedean copulas which we described in the Section IV.1.10.

The quality of fit can easily be compared by checking the error term $\varepsilon(\Theta)$. To our surprise the Gumbel copula provides completely inaccurate, off-market prices. The Clayton copula yields better results, roughly comparable to that of the normal (Gaussian) copula. This is not a new discovery, similar finding is reported e.g. by (Brigo, 2010). These results

¹⁴²if the upfront is quoted, it is first necessary to convert the upfront quote to the fair spread quote against which the model spread is compared. This is done in the following way. Firstly, we find ρ in the Gaussian copula model such that $u_{\text{model}}(\rho) = u_{\text{market}}$, where $u_{\text{model}}(\rho) = C_{[a,d]} \left(1 - \frac{s^*}{s(\rho)} \right)$. Then we use the correlation ρ for calculation of $s = s(\rho)$ which represents the upfront 'market quote' expressed as a fair spread. This quantity s then enters the least-squares problem.

¹⁴³clearly, this can not be evaluated in the case of the Archimedean copulas, where there is no correlation parameter.

are not surprising and could further be improved. For example, (Choroś-Tomczyk et al., 2014) applies an econometric analysis to make the copula parameter θ a function of time. As an alternative, (Okhrin et al., 2013) proposes hierarchical Archimedean copulas which allow more complex structures to be fit.

In our study the NIG and double t copulas provide both much better fit than the Gaussian copula, in almost all cases the total error produced by these two copulas is only about 5% of the error produced by the Gaussian copula. Significant improvement over the basic Gaussian copula model is also the Normal+ model with the stochastic correlation. Arguably the best fit is achieved with the double $t+$ model. A particular feature we found out in the case of the double $t+$ model is that the best fit is typically achieved with ρ close to 1 and ν_X close to 2 such that the fit to the 'current' market quotes is controlled rather through ν_Y than through the correlation ρ .

From the Tables 16–19 it is furthermore easy to see that in general the models except for the Gaussian, Gumbel and Clayton copulas manage to fit the junior tranches but some have difficulties with fitting the super senior tranches. In this context, the double $t+$ model provides the most consistent results although the (super)-senior tranches are still relatively far from being perfectly fit.¹⁴⁴ However, not only the double $t+$ model provides the best fit, yet it also can produce (almost) monotonous tranche spreads with respect to the change in ν_Y . An example is shown in Figure 29, where only the equity tranche gently loses its monotony at $\nu_Y > 4$, otherwise the spreads are monotonous with the change of ν_Y . Also notice that ν_Y has the major effect on the tranche spreads for values between 2 and 3. Based on our analysis, the double $t+$ copula model is therefore the most attractive models from those included in the comparison. It seems that in nearly all the pricing cases the values (ρ, ν_X) in the double $t+$ model are given to be close to 1 and 2, respectively and reflect the general pattern of the credit market, while the third parameter ν_Y serves to fine-tune the fit.

¹⁴⁴this can be still circumvented. See our concluding remarks.

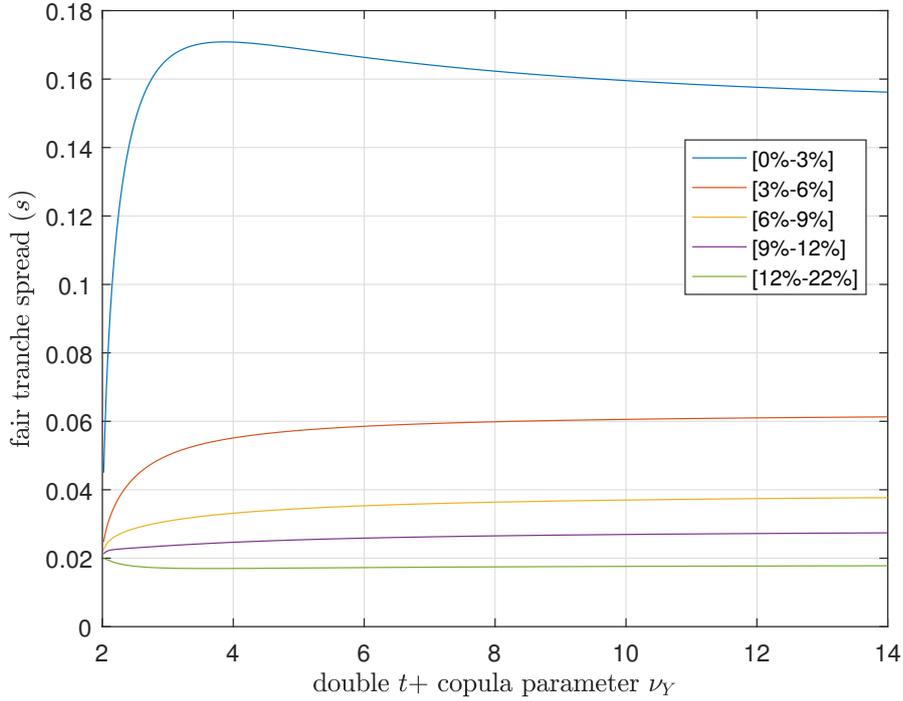


Figure 29: Curves of the fair tranche spreads plotted as functions of ν_Y produced by the double $t+$ copula calibrated to 2013 data with (ρ, ν_X) being fixed.

quote	normal	normal+	NIG	NIG+	double t	double $t+$	Clayton	Gumbel
0.2066	0.1985	0.2063	0.2074	0.2074	0.2075	0.2071	0.1975	0.1906
0.0609	0.0905	0.0633	0.0593	0.0587	0.0588	0.0608	0.0943	0.1121
0.0353	0.0572	0.0313	0.0333	0.0339	0.0337	0.0334	0.0594	0.0781
0.0188	0.0394	0.0223	0.0243	0.0245	0.0246	0.0238	0.0404	0.0559
0.0091	0.0203	0.0211	0.0168	0.0163	0.0168	0.0158	0.0199	0.0266
$\varepsilon(\Theta)$	19.72	1.80	0.97	0.93	1.01	0.75	23.63	63.91
ρ	0.4418	0.3453	0.5775	0.5817	0.5864	0.9946	N/A	N/A

Table 16: The 2008 dataset: quoted and fitted model spreads.

IV.1.12 Concluding remarks

In this part of the thesis we were looking for an optimal bottom-up model for the CDO pricing. As a benchmark, we used the traditional, yet for many reasons deficient, Gaussian copula. For the purpose of the benchmarking we collected 4 full datasets of 5-year quotes of tranches of on-the-run CDO iTraxx index. The models we investigated were NIG copula and its extension NIG+, double t and its extension double $t+$, a Gaussian copula with stochastic correlation (denoted 'Normal+') and two Archimedean copulas: Clayton and Gumbel copulas. The Clayton and Gumbel copulas did not generate convincing results. While the Clayton copula produced at least prices of accuracy comparable

quote	normal	normal+	NIG	NIG+	double t	double $t+$	Clayton	Gumbel
0.1584	0.1522	0.1581	0.1588	0.1588	0.1589	0.1584	0.1512	0.1447
0.0594	0.0769	0.0611	0.0591	0.0590	0.0589	0.0609	0.0798	0.0946
0.0404	0.0522	0.0381	0.0380	0.0381	0.0382	0.0372	0.0544	0.0717
0.0260	0.0384	0.0262	0.0286	0.0284	0.0288	0.0269	0.0397	0.0563
0.0137	0.0224	0.0185	0.0194	0.0188	0.0193	0.0174	0.0225	0.0340
$\varepsilon(\Theta)$	7.12	0.31	0.44	0.37	0.44	0.27	9.26	37.34
ρ	0.5508	0.5666	0.6334	0.6399	0.6346	0.9920	N/A	N/A

Table 17: The 2013 dataset: quoted and fitted model spreads.

quote	normal	normal+	NIG	NIG+	double t	double $t+$	Clayton	Gumbel
0.0891	0.0842	0.0892	0.0894	0.0894	0.0894	0.0893	0.0836	0.0793
0.0295	0.0441	0.0299	0.0287	0.0288	0.0286	0.0290	0.0457	0.0555
0.0153	0.0271	0.0155	0.0176	0.0174	0.0178	0.0173	0.0282	0.0405
0.0035	0.0055	0.0077	0.0071	0.0072	0.0071	0.0065	0.0052	0.0092
$\varepsilon(\Theta)$	3.78	0.17	0.19	0.18	0.20	0.13	4.64	14.37
ρ	0.6497	0.5686	0.7246	0.7319	0.7238	0.9988	N/A	N/A

Table 18: The 2014 dataset: quoted and fitted model spreads.

quote	normal	normal+	NIG	NIG+	double t	double $t+$	Clayton	Gumbel
0.1383	0.1285	0.1385	0.1387	0.1387	0.1385	0.1385	0.1277	0.1197
0.0331	0.0640	0.0332	0.0319	0.0320	0.0335	0.0333	0.0663	0.0785
0.0147	0.0372	0.0174	0.0185	0.0183	0.0194	0.0192	0.0385	0.0531
0.0047	0.0063	0.0099	0.0095	0.0095	0.0091	0.0089	0.0059	0.0086
$\varepsilon(\Theta)$	15.56	0.34	0.39	0.38	0.41	0.37	17.83	38.87
ρ	0.5561	0.3142	0.6861	0.6917	0.7173	0.6307	N/A	N/A

Table 19: The 2015 dataset: quoted and fitted model spreads.

with the Gaussian copula, the Gumbel copula failed completely to match the market. The NIG, double t and the stochastic correlation approaches produced much better and consistent results. We found out that the total error of these models was approximately only 5% of the error generated with the Gaussian copula. The double t and NIG models produced very similar and small errors although they are only 2-parameter models.

The best result in all cases were, however, achieved with the three-parameter double $t+$ model where the smallest error was produced. We also found an interesting property that two parameters of the three in the model attained in all the calibration cases very similar values and only the third parameter ν_Y has differed. This could point to some sort of optimality of this choice capturing the general structure of the credit market with the two parameters and adapting the model to the current market situation by the third parameter. None of the models, however, was able to fit perfectly the *whole* structure of the market quotes as it is particularly difficult to match the exact prices of the (super)-senior tranches. Despite this imperfection, if one is about to price bespoke tranches that are not (super)-senior, double $t(+)$ or NIG($+$) models will do much better than the Gaussian copula. To price the (super)-senior tranches, we recommend the following approach based on the stability of the parameters (ρ, ν_X) in the double $t+$ copula. First, calibrate the model to *all* tranche quotes in the dataset. Notice the first two parameters (ρ, ν_X) proved to be stable quantities. Once this is done, fix the parameters (ρ, ν_X) and recalibrate the third parameter ν_Y to match the market quotes of the (super)-senior tranches. After this is done the model will provide off-market prices for the junior tranches yet accurate prices of the possibly bespoke (super)-senior tranches.

IV.2 Modeling the portfolio risk

The problem of modeling the portfolio risk is obviously one of the most challenging tasks in finance. The main concern with the portfolio modeling is that of the interlinkages between defaults, commonly referred to as the 'default correlation' or 'credit correlation'. This is because banks need to hold capital buffers against the unexpected losses (ULs) of their portfolios but this UL is a heavily dependent quantity on the credit correlation, given the probabilities of default (PDs) and the recoveries are estimated correctly. In general, however, the estimation of PDs and recoveries is thought to be less difficult than the specification of credit correlations as PDs and recoveries follow from the observed events but correlations are rather latent. Last but not least for the credit portfolio modeling, when compared to the market risk modeling, there is usually not that much data to be used by the model. Hence, the inputs are relatively limited and possibly noisy, which means that mostly the banks are just using PDs and recoveries.

In practice, two classes of approaches usually appear in the credit portfolio modeling: the asset-value models and the reduced-form models. This does therefore not differ from the obligor-level credit models which we discussed in a great detail in the Part II and III, respectively. The reduced-form models typically offer a better analytical tractability but lack a powerful framework for incorporation of correlations (see the discussion in the

Section III.2.2.1). The asset-value models, on the other hand, usually allow for a much more effective and explicit treatment of correlations but lack the analytical tractability of the reduced-form models.

Our intention is to combine the strengths of both the model classes above and construct a complex model *framework*, that is to adopt the analytical tractability of the reduced-form models and allow for the correlations from the asset-value models. Also, the default rates might not be considered as fixed numbers but rather uncertain quantities. Another issue is that of recoveries experienced after debtor's default. In the simplest possible case, recoveries are assumed to be a fixed portion of the defaulted exposure. In more advanced modeling approaches, recoveries (or loss-given defaults, complementarily) follow some distribution and thus are regarded as being stochastic. We allow for all these features to be captured by the model.

With an emphasis on applicability, we should also not overlook the approach of the regulator. While Basel I (BCBS, 1988, 1996) assumed away any deeper characteristics of the credit portfolios (such as portfolio correlations), In Basel II (BCBS, 2004, 2006) and further the internal rating-based (IRB) approach has been proposed to calculate more sensitively the *regulatory* capital needed to cover the unexpected losses. Even that this IRB approach accounts for some portfolio particularities, this is still insufficient for many portfolios. For example, the recoveries might worsen during the crisis and this possibility is not captured by the IRB formula. Also the IRB formula is derived from Vasicek's large homogeneous credit pool (LHP) formula. Hence, the IRB formula is based on LHP, yet in the real world there is no such a LHP portfolio and the portfolios are often completely non-homogenous. In spite of these imperfections, we should not be contradiction with the regulator. We thus propose a framework that is in general *compatible with IRB*, yet is much more customizable to account for the portfolio-specific features. We also show that in the limit, the framework produces outputs similar to Vasicek's LHP. We find the IRB-compatibility useful as the model used for internal capital should not be based on completely different assumptions than a (IRB) model used for the regulatory capital.

The recent studies of (Lall, 2012; Witzany, 2013) accentuate that the regulatory formula (based on Gaussian copula) underestimates the credit risk. (Witzany, 2013) comes with the conclusion that the capital held against the credit risk should be approximately twice the capital computed using the Gaussian copula regulatory models. We show, that the double t model, when used in this proposed framework, produces UL very close to the double of the Gaussian copula-based UL given all the inputs are the same. In the widely recognized papers of (Gordy, 2003; Pykhtin, 2004) the authors also work with a quantity called *granularity adjustment* (GA) that should be accounted for in the models as the IRB-based models are based on a stylized infinitely granular (numerous) portfolio yet the real portfolios are not infinitely granular. In our approach, the GA is an inherent feature of the model as we work with a finite portfolio directly and do not derive finite portfolio from the infinite one.

IV.2.1 Generator of default rates

The most critical part of the portfolio credit modeling is the generation of default events. Similarly to the basket credit derivatives, we assume the probability of the debtor i to default is subject to a factor function (IV.1.3) already known from the CDO modeling

$$Z_i = \sqrt{\rho_i} m_X X + \sqrt{1 - \rho_i} m_{Y_i} Y_i,$$

that is, we consider the portfolio of credits to be linked to a systematic factor X . Although this specification can be limiting,¹⁴⁵ it is, in general, very hard to work with more sophisticated correlation structure because the estimation of default correlations is very challenging in practice. Notice that this specification also is consistent with the IRB formula if X and Y_i are normal. Let q_i be the *statistical* probability of default of the credit i . The $X = x$ -conditional probability of default equals

$$Q_i(x) = F_{Y_i} \left(\frac{F_{Z_i}^{-1}(q_i) - \sqrt{\rho_i} m_X x}{\sqrt{1 - \rho_i} m_{Y_i}} \right).$$

This is the 'deterministic' variant as the statistical probability q_i is given and so the function Q_i does not differ from that of the CDO specification. We now allow the statistical default probability q_i to be a random variable itself, with mean μ_i and standard deviation σ_i . Suppose F_{q_i} is a distribution of q_i and let $h > 0$ be a reasonably small number such that

$$q_i(j) = \Pr[q_i = jh] = F_{q_i} \left(\left(j + \frac{1}{2} \right) h \right) - F_{q_i} \left(\left(j - \frac{1}{2} \right) h \right), j = 1, \dots, J; \left(J + \frac{1}{2} \right) h \leq 1.$$

Then

$$Q_i(x) = \sum_{j=1}^J q_i(j) F_{Y_i} \left(\frac{F_{Z_i}^{-1}(jh) - \sqrt{\rho_i} m_X x}{\sqrt{1 - \rho_i} m_{Y_i}} \right),$$

where $Q_i(x)$ now has the meaning of expected PD, conditional on $X = x$. Similarly to (CreditSuisse, 1997), we make a further assumption about q_i that it has a Gamma distribution with mean μ_i and a variance σ_i^2 , which in turn means that $q_i \sim \Gamma \left(\frac{\mu_i^2}{\sigma_i^2}, \frac{\sigma_i^2}{\mu_i} \right)$. Surprisingly, in our model, this 'stochastic PD' adjustment *decreases* the unexpected loss populated by the model. In the Figure 30 we show the change in the distribution of number of defaults when the stochasticity of PDs is taken into account. In CreditRisk⁺ the UL increases due to the PD uncertainty because the main assumption there is that conditional on some common factor, the actual PDs tend to move the same direction, while here the uncertainty in the unconditional PDs is credit-specific and the simultaneous defaults are determined by the factor function. The fact that in our case the UL actually decreases with increased PD-uncertainty is the best illustrated in the following example. It is a known fact that the tail of the distribution is mostly caused by the simultaneous losses occurring when X is low. Hence, it makes sense to focus only on what happens when X is low. When we have the fixed PD model, then for some low value of X the debtor will 'almost surely' default. Given the same fixed X at low value, it is still possible that the

¹⁴⁵instead of having the correlation matrix specified explicitly such as in CreditMetrics (see (Gupton et al., 1997)), it is given implicitly such that assets i, j have the correlation $\sqrt{\rho_i \rho_j}$.

credit does not default as the default thresholds are distributed according to the chosen distribution of the PDs. This, in turn, implies some sort of 'diversification' allowing for no-default when X is very low and on the other hand make the default possible when X is very high (this, however, does not contribute to the tail loss). The technical reason why this actually occurs is that the distribution of PD is independent of X . We present this stochastic PD adjustment rather for completeness but it will not be assumed later in the text.

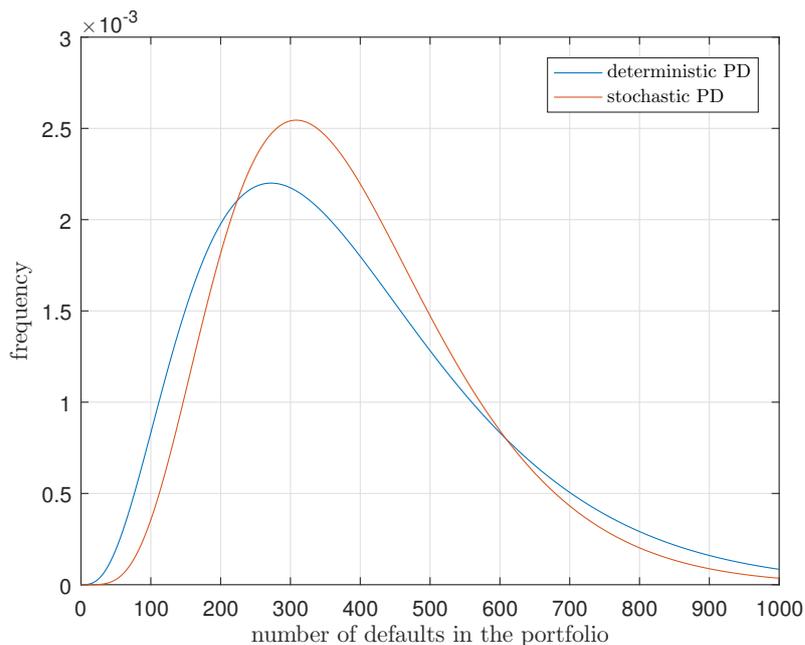


Figure 30: The effect of the stochastic default rates on the number of modeled defaults in the portfolio.

IV.2.2 Generating the distribution of number of defaults

In the previous section it was described how the model generates the default probabilities, conditional on $X = x$. These probabilities can be subsequently combined to generate the full distribution of the number of defaults in the portfolio. To make so, the model uses the probability-generating-functions (PGFs) P_i . Let i th debtor's conditional default PGF be

$$P_i(z|x) = 1 - Q_i(x) + Q_i(x)z.$$

Then, the PGF of number of defaults in the portfolio P , is given by

$$P(z) = \int_D f_X(x)P(z|x)dx = \int_D f_X(x) \prod_{i=1}^n P_i(z|x)dx,$$

where D is the support of f_X , as usual. An efficient way of computing P is to apply the multidimensional FFT (and IFFT) for the computation of the product $P(z|x)$ at various values of x . For the integration we suggest to use the Gauss-Legendre quadrature.

IV.2.3 Mapping of default losses to the default events

As the main concern is the estimation of losses, it is necessary to map the default events to losses and we, in general, assume that the severities of the losses also occur with some probabilities conditional on $X = x$, that is LGD/RR is a stochastic quantity. This can be incorporated in the following way. Denote $\Pr [\delta_i = \delta_j^* | x]$ the $X = x$ -conditional probability that i th debtor's recovery δ_i after default attains some value $\delta_j^* \in [0, 1], j = 0, \dots, J$. Denote E_i the notional exposure of the credit i and let v be a unit of accounting.¹⁴⁶ Furthermore, let

$$c(E_i, \delta_j^*) = \left\lceil \frac{E_i(1 - \delta_j^*)}{v} \right\rceil,$$

be a function mapping the loss to the nearest equal or greater integer value. Then the $X = x$ -conditional loss-generating function $G_i(z|x)$ of a debtor i has the form

$$G_i(z|x) = 1 - Q_i(x) + Q_i(x) \sum_{j=0}^J \Pr [\delta_i = \delta_j^* | x] z^{c(E_i, \delta_j^*)}. \quad (\text{IV.2.1})$$

The unconditional PGF $G(z)$ representing the losses over the whole portfolio is then obtained the same way as the PGF P , that is

$$G(z) = \int_D f_X(x) G(z|x) dx = \int_D f_X(x) \prod_{i=1}^n G_i(z|x) dx,$$

and it can be calculated by applying the same techniques as for P .

IV.2.4 Deterministic and stochastic recoveries

For many many real-world scenarios, banks have estimates of how the recoveries could vary rather than having one single point estimate of the RR (see e.g. (Kolman et al., 2014) the estimation of RR using the survival analysis). Often also, it is nearly impossible to determine RR in advance but statistically the bank could have a collection of historical observations of recoveries. We push these ideas further so that the model structure is adjusted to allow for distributional RR specification of every credit in the basket.

The framework can work with deterministic (fixed), or stochastic RR. The deterministic case is trivial. Given a predefined set of recoveries $\delta_j^*, j = 0, \dots, J$ and their explicitly stated probabilities $\Pr [\delta_i = \delta_j^*]$, then we just replace $\Pr [\delta_i = \delta_j^* | x]$ with $\Pr [\delta_i = \delta_j^*]$ in (IV.2.1). In the most cases, RR/LGD of debtor i is assumed to be just one number δ^* , not a set of numbers. In such a case the PGF (IV.2.1) reduces just to

$$G_i(z|x) = 1 - Q_i(x) + Q_i(x) z^{c(E_i, \delta^*)},$$

as this is a special case of the deterministic case with $\Pr [\delta_i = \delta^*] = 1$.

¹⁴⁶this unit has the importance for expressing the exposures in multiples of v . If in a portfolio the typical exposure is, say, in the order of millions, it would be a good idea to choose $v = 1000000$ such that each exposure expressed in v is a relatively small integer value.

In a more general and realistic setup, recoveries can be stochastic and generated from some distribution. Typically, in the retail segment of unsecured loans, recoveries exhibit a 'U-shape' (or 'bathtub' shape) (see e.g. the research articles of (Zhang and Thomas, 2012) or (Kolman et al., 2014)) which means high and low recoveries occur quite frequently while the average recoveries are less common. In the corporate segment the recoveries depend on the collateral and seniority of the debt and so the shape of the recovery distribution is mainly determined by these factors. We thus adopt the idea that each debt i , be it retail or corporate, can have its own distribution of the recovery rate although in practice there will be only a few recovery distributions and each debt will be linked to some of these. We further assume that the *recoveries are correlated* with the systematic factor X (see (Witzany, 2009) for an example of the estimation procedure, or (Miu and Ozdemir, 2006) who argue that LGD have to increase dramatically if one is about to relax their correlations yet he wants to maintain the same level of risk). This in turn means that recoveries are also implicitly correlated with PDs. For the purpose of this text, we assume the recovery of i th debtor δ_i *approximately* stems from the Beta(α_i, β_i) distribution (for the motivation for using the beta distribution in the RR/LGD modeling and its properties, see for example (Schönbucher, 2003, Section 6)), although any distribution of recoveries best fitting to the data sample can be chosen. The next main issue is the attribution of the recovery distribution to the market scenario determined by the realization of X . Let us assume that we have cut the Beta distribution into 'recovery cohorts' determined by the recovery thresholds $b_j, j = 0, \dots, J$ such that

$$\delta_i = \begin{cases} \delta_0^*, & \text{if } B_i \leq b_0 \\ \delta_1^*, & \text{if } b_0 < B_i \leq b_1 \\ \vdots & \\ \delta_J^*, & \text{if } b_{J-1} < B_i \leq b_J \end{cases}, B_i \sim \text{Beta}(\alpha_i, \beta_i).$$

The distance of the thresholds $(b_j - b_{j-1})$, i.e. the size of the j th cohort depends on the available data, and the spacing does not necessarily need to be equidistant. The assigned δ_j^* should be somewhere between b_{j-1} and b_j , for example in the midpoint. Assume the recoveries are driven by the factor function

$$Z_i^\delta = \sqrt{\rho_i^\delta} m_X X + \sqrt{1 - \rho_i^\delta} m_{Y_i} Y_i^\delta,$$

where ρ_i^δ determines how closely the recovery distribution of i th debtor is linked to the common market factor X . We state that recovery $\delta_i = \delta_0^*$ if,

$$\sqrt{\rho_i^\delta} m_X X + \sqrt{1 - \rho_i^\delta} m_{Y_i} Y_i^\delta \leq F_{Z_i^\delta}^{-1}(F_{B_i}(b_0)).$$

Hence, conditional on $X = x$ we have

$$\Pr[\delta_i = \delta_0^* | x] = F_{Y_i^\delta} \left(\frac{F_{Z_i^\delta}^{-1}(F_{B_i}(b_0)) - \sqrt{\rho_i^\delta} m_X x}{\sqrt{1 - \rho_i^\delta} m_{Y_i}} \right).$$

For other cohorts, observe that

$$\Pr[\delta_i \leq \delta_j^* | x] = F_{Y_i^\delta} \left(\frac{F_{Z_i^\delta}^{-1}(F_{B_i}(b_j)) - \sqrt{\rho_i^\delta} m_X x}{\sqrt{1 - \rho_i^\delta} m_{Y_i}} \right), 1 \leq j \leq J.$$

This suggests for $j \geq 1$

$$\begin{aligned} \Pr [\delta_i = \delta_j^* | x] &= \Pr [\delta_i \leq \delta_j^* | x] - \Pr [\delta_i \leq \delta_{j-1}^* | x] \\ &= \Pr [\delta_i \leq \delta_j^* | x] - \sum_{k=0}^{j-1} \Pr [\delta_i = \delta_k^* | x]. \end{aligned}$$

To this end we shall note that some texts (for example (Ou et al., 2016, p. 7)) argue that RR is negatively correlated with PD. In such a case, the factor function we use for the RR can not be used as it expects $\rho_i^\delta \geq 0$. It can, however, be easily circumvented by using LGD (instead of RR) in the factor function, such that some $\rho_i^{\text{LGD}} \geq 0$ would be used. If then X is low (PDs are high), LGD would be low implying to a certain degree negative correlation between PDs and RR.

IV.2.5 The double t specification

So far we have been silently assuming the Normal model with (X, Y_i, Z_i) stochastic. The application of the Normal model is, however, questionable as it is known that the Gaussian copula does not fully capture the nature of the in-crisis credit markets (see for example (Jones, 2009) or (Salmon, 2012)). It could thus potentially underestimate the credit VaR and this implies less capital than needed. As the double t copula has proved strong in the modeling of the credit products, it could be potentially interesting to use it also for the portfolio risk modeling. Since this framework has a general structure, it can be adjusted to work with the double t copula with $\nu_X = \nu_Y = \nu$ too. This, according to the Section IV.1.8 means to set

$$Z_i = \sqrt{\rho_i} \sqrt{\frac{\nu-2}{\nu}} X + \sqrt{1-\rho_i} \sqrt{\frac{\nu-2}{\nu}} Y_i \text{ with } X, Y_i \sim t_\nu, \quad (\text{IV.2.2})$$

where $\sqrt{\frac{\nu-2}{\nu}}$ is the standardizing constant. If, furthermore, the recovery rate is stochastic, then the factor function for recovery δ_i of i th debtor reads

$$Z_i^\delta = \sqrt{\rho_i^\delta} \sqrt{\frac{\nu-2}{\nu}} X + \sqrt{1-\rho_i^\delta} \sqrt{\frac{\nu-2}{\nu}} Y_i^\delta \text{ with } Y_i^\delta \sim t_\nu,$$

that is, we are forced to adopt the double t copula also for the recovery.¹⁴⁷ The other building blocks of the model remain unchanged. The main complication with the double t copula is that it is computationally demanding. For special cases of ν we have shown in (Kolman, 2014a) that solutions for the double t copula are trivial, only in terms of elementary functions. This is supported by (IV.1.8) where some sort of series representation is shown for the modified Bessel function of the third kind $K_{\frac{\nu}{2}}$ with ν being odd. In fact for the important choice $\nu = 3$ we derived in (Kolman, 2014a) that indeed a nice formula for the CDF F_{Z_i} exists.

¹⁴⁷this specification, however, does not hold absolutely. We are in any case bound to adopt the fact that X will always be t -distributed but Y_i^δ can be specified such that it has a t distribution with different degrees of freedom ν_Y than X . If we in turn let $\nu_Y \rightarrow \infty$ then Y_i is effectively Normal but we can still use the formulas for double t copula that we derived earlier in the Section IV.1.8.

Theorem IV.2.1 (F_{Z_i} in double t with $\nu = 3$)

Assume the factor Z_i as in (IV.2.2). Then for $\nu = 3$ the CDF of Z_i has a completely analytical form given as follows. Set

$$r = 1 + 2\sqrt{(1 - \rho_i)\rho_i} \text{ and } p = \frac{1}{12\pi r^5}.$$

Then

$$\begin{aligned} F_{Z_i}(z) = & p [aC(15, 7, 1, 7, -15, -55, -73, 15, z|r) + bC(9, 5, 3, 6, -3, -11, 11, 3, z|r)] \\ & + p [cC(15, 3, 1, 5, -15, 73, 55, 15, z|r) + dC(105, 1, 1, 4, 279, 511, 385, 105, z|r)], \end{aligned}$$

where

$$\begin{aligned} a &= \sqrt{1 - \rho_i} - \rho_i \sqrt{1 - \rho_i} + \rho_i \sqrt{\rho_i} \\ b &= 3\sqrt{1 - \rho_i} + 10\sqrt{\rho_i} + 9\rho_i \sqrt{1 - \rho_i} - 5\rho_i \sqrt{\rho_i} - 2\rho_i^2 \sqrt{1 - \rho_i} - 2\rho_i^2 \sqrt{\rho_i} \\ c &= 3\sqrt{1 - \rho_i} + 20\sqrt{\rho_i} + 49\rho_i \sqrt{1 - \rho_i} + 25\rho_i \sqrt{\rho_i} \\ &\quad - 12\rho_i^2 \sqrt{1 - \rho_i} - 72\rho_i^2 \sqrt{\rho_i} - 20\rho_i^3 \sqrt{1 - \rho_i} + 20\rho_i^3 \sqrt{\rho_i} \\ d &= \sqrt{1 - \rho_i} + 10\sqrt{\rho_i} + 39\rho_i \sqrt{1 - \rho_i} + 65\rho_i \sqrt{\rho_i} \\ &\quad + 38\rho_i^2 \sqrt{1 - \rho_i} - 94\rho_i^2 \sqrt{\rho_i} - 92\rho_i^3 \sqrt{1 - \rho_i} - 4\rho_i^3 \sqrt{\rho_i} + 24\rho_i^4 \sqrt{1 - \rho_i} + 24\rho_i^4 \sqrt{\rho_i}, \end{aligned}$$

and

$$\begin{aligned} C(c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, z|r) &= \frac{c_1}{32} \pi r^{\frac{c_2}{2}} + \frac{c_3}{16} z^{c_2} (A(\cdot|r) + B(\cdot|r)) \\ A(c_2, c_4, c_5, c_6, c_7, c_8, z|r) &= \frac{r^{c_4} (c_5 + \frac{c_6}{r} z^2 + \frac{c_7}{r^2} z^4 + \frac{c_8}{r^3} z^6)}{z^{c_2-1} (r + z^2)^4} \\ B(c_2, c_8, z|r) &= \frac{c_8 r^{\frac{c_2}{2}} \operatorname{arccot}\left(\frac{\sqrt{r}}{|z|}\right)}{|z|^{c_2}}. \end{aligned}$$

This expression is valid of $F_{Z_i}(z), z \neq 0$. Finally, $F_{Z_i}(0) = \frac{1}{2}$ by definition of a symmetrical distribution.

Remarks. (i) The Theorem IV.2.1 allows for a very quick computation of $F_{Z_i}, \nu = 3$, yet for the computations in the model we need $F_{Z_i}^{-1}$. The formula for F_{Z_i} presented here can further be even analytically inverted yet the inverted form is too lengthy to be presented here. (ii) Analogous form can also be derived for $F_{Z_i}, \nu = 5, 7, \dots$, yet we find the case with $\nu = 3$ the most useful and relevant one and thus other forms are omitted.

IV.2.6 The convergence to the Vasicek's LHP

Oldrich Vasicek (Vasicek, 1991, 2002) has derived a famous limiting loss distribution for loss of infinitely granular portfolio of a completely homogeneous debt.¹⁴⁸ In Vasicek's view, all the credits are subject to the same probability of default $q_i \equiv q$ and the same correlation $\rho_i = \rho$ with the common factor X . Vasicek found that

$$F(p) = \Pr[\text{relative number of defaults} \leq p] = N\left(\frac{\sqrt{1 - \rho}N^{-1}(p) - N^{-1}(q)}{\sqrt{\rho}}\right),$$

¹⁴⁸this is the concept of the 'large homogeneous portfolio' (LHP).

for $p \in [0, 1]$. This can be easily formulated for the relative *loss* which differs from the relative number of defaults by the fact that recovery δ is involved. If each defaulted debt recovers $\delta = 0$, then LGD is $(1 - \delta) = 1$ and the expression above also holds for the relative loss in the portfolio. If $\delta > 0$, then the portfolio loss expression changes slightly. If all credits default, this corresponds to $F(p), p = 1$ and the overall relative loss with respect to the notional equals $(1 - \delta)$. Hence, the probability $F(p)$ corresponds to the probability of losing $\tilde{p} = (1 - \delta)p$ of the portfolio notional. Let $\tilde{F}(\tilde{p}) = \Pr[\text{relative loss} \leq \tilde{p}]$, $\tilde{p} \in [0, 1 - \delta]$ and we have

$$\tilde{F}(\tilde{p}) = F\left(\frac{\tilde{p}}{1 - \delta}\right) = N\left(\frac{\sqrt{1 - \rho}N^{-1}\left(\frac{\tilde{p}}{1 - \delta}\right) - N^{-1}(q)}{\sqrt{\rho}}\right),$$

which captures the idea of mapping the losses \tilde{p} to p by considering how many relative defaults p have to occur in order to cause $\tilde{p} = p(1 - \delta)$ relative portfolio loss.¹⁴⁹ The issue here, however, is to compare PDFs not the CDFs of the loss distributions. A tedious calculation shows that taking $f(p) = dF(p)/dp$ gives (less known) equations for the PDFs

$$f(p) = \sqrt{\frac{1}{\rho} - 1} \exp\left(\frac{1}{2}[N^{-1}(p)]^2 - \frac{[N^{-1}(q) - \sqrt{1 - \rho}N^{-1}(p)]^2}{2\rho}\right) \Rightarrow \tilde{f}(\tilde{p}) = f\left(\frac{\tilde{p}}{1 - \delta}\right),$$

where f is the PDF of the relative number of defaults and \tilde{f} is the PDF of the relative loss of the LHP. In Figure 31 we show how the proposed model converges to this Vasicek's LHP for $n \rightarrow \infty$ by comparing the loss densities (frequencies). It is also to see that the less numerous portfolio *ceteris paribus* generates fatter tails. This was indeed the motivation behind introducing the granularity adjustment (GA) discussed in (Gordy, 2003; Pykhtin, 2004). Since we take the portfolio as it is, in our framework we always generate the precise loss distribution and so GA is an inbuilt feature of the framework and does not require any further treatment.

IV.2.7 Speed issues and recurrence relations

If the portfolio is small, then it is computationally simple to calculate the loss distribution without any necessary adjustments. However, with the increasing portfolio size, the computational complexity grows exponentially. The computational intensity is governed by several model components, each having a completely different computational pattern.

Component 1: Generating the $X = x$ -conditional probabilities of default $Q_i(x)$. The main cost here is the necessity to calculate $F_{Z_i}^{-1}(q_i)$ for every i , where $F_{Z_i}^{-1}$ might depend on ρ_i . Furthermore, when the volatility of the default rates is assumed then multiple evaluations of $F_{Z_i}^{-1}$ are needed as it must be evaluated for every possible value of the (stochastic) default probability. If then, for example, the double t copula specification of the factor function is used, this can consume a significant portion of time. The good thing is that $F_{Z_i}^{-1}$ can be precomputed and used all over for the same

¹⁴⁹For example if we look for the probability of having lost less than 0.1 of the portfolio notional, given $\delta = 0.4$, this is the same problem as finding the probability of relatively defaulting p which is given as $p = \frac{\tilde{p}}{1 - \delta} = \frac{0.1}{0.6} = 0.1667$ (causing the relative loss of 0.1).

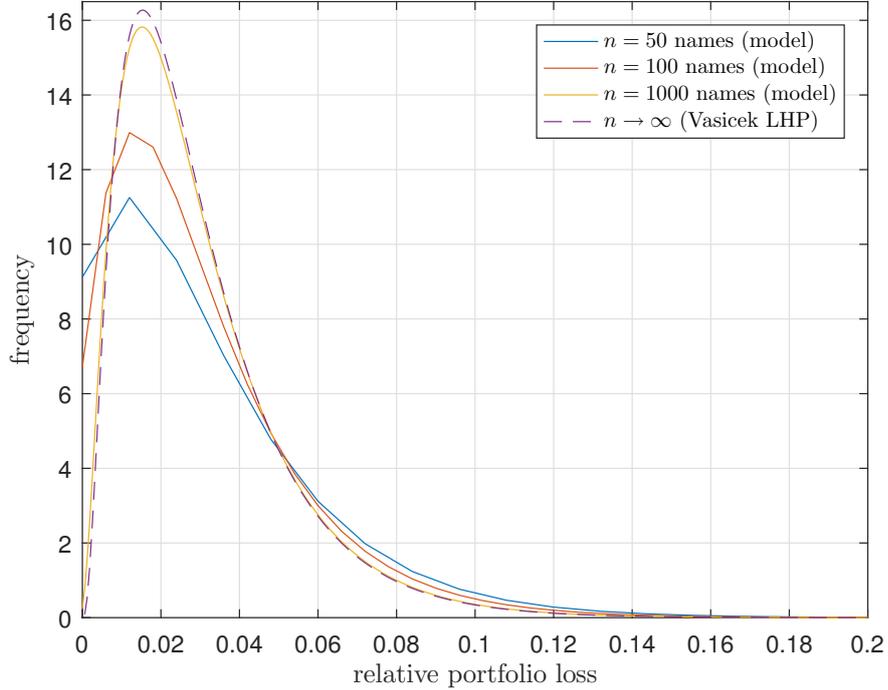


Figure 31: Model convergence to Vasicek LHP, for $n = 50, 100, 1000$. $q = 0.05, \rho = 0.1, \delta = 0.4$.

name. Another point is that $Q_i(x)$ has to be evaluated at all values of $X = x$. In the Gaussian setup (X, Y_i, Z_i) are normal and $Q_i(x)$ can be simultaneously computed for all i and all values of x . In the double t specification, we can do this only partially (evaluate at all $X = x$). In general, this part of the model is not the most time-consuming part.

Component 2: The common factor X . In order to obtain the unconditional distribution of losses, firstly the $X = x$ -conditional probabilities of default have to be computed. Clearly, the more points of x , the lengthier is the calculation, yet only linearly. In the Gaussian setup, we found $x_{\min} = -4.3, x_{\max} = 4.3$ with 100 integration points (Gauss-Legendre quadrature which can be fully vectorized) as optimal. The use of less points causes unnatural groupings of high and low probabilities on the probability distribution of losses and biases the distributional properties. Comparison of an insufficient only 40-point quadrature and the suggested 100-point quadrature is illustrated in the Figure 32.

Component 3: The aggregation of losses in the portfolio from the individual $X = x$ -conditional loss-PGFs $G_i(z|x)$. This part can really bottleneck the whole calculation procedure. There are two sources of the computational difficulty: (i) numerous portfolio, (ii) portfolio with very large exposures, or combination of both. In a very numerous portfolio, there are many $G_i(z|x)$ to be aggregated. In a portfolio of very large exposures, $G_i(z|x)$ contain polynomials of a very high order which slow down the convolutions. There are several optimization possibilities for the Component 3. The simplest and the most intu-

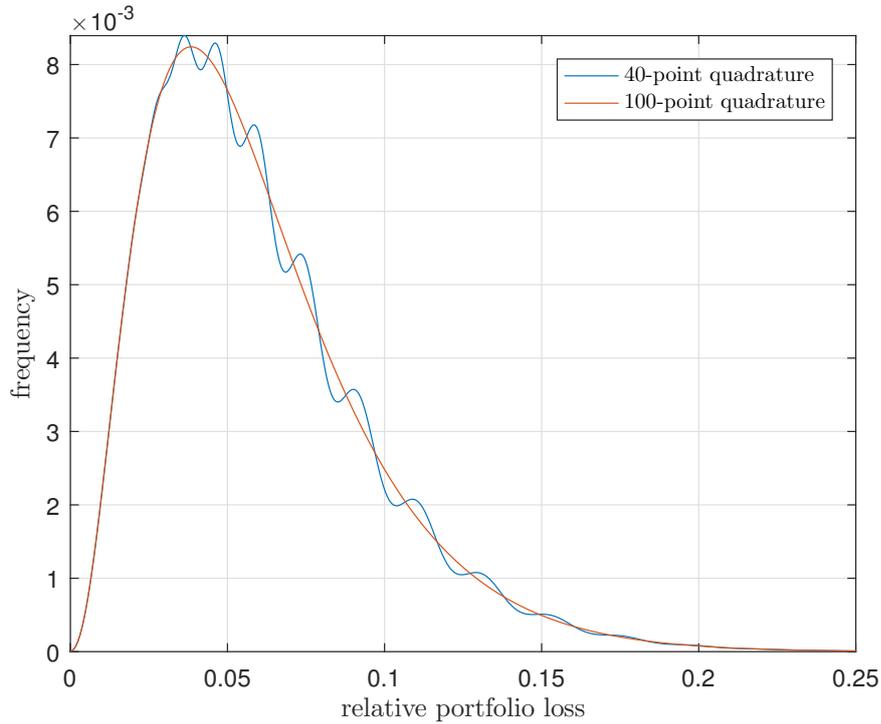


Figure 32: Loss probability distribution under a sufficient and insufficient integration.

itive step is to increase the unit size v which furthermore can be combined with any other step presented here. This reduces the order of the polynomials in $G_i(z|x)$ which are then easier to be convoluted. It, however, comes at the cost of precision.¹⁵⁰ If one uses a convolution algorithm to compute the aggregated losses and assumes a non-stochastic RR/LGD for every debt, then the *reduced convolution* algorithm that we developed uniquely for this type of application and is presented in the Appendix B might bring a significant speed gain.

The most efficient way of the polynomial multiplication, however, is the FFT approach. Although this outperforms the convolution in speed it still slows down significantly when the portfolio is large. We thus propose a specialized optimization of the FFT method which we call the **clustered FFT**. The idea is to break the portfolio down to smaller units of risky exposures (clusters), apply FFT and IFFT on the exposures in the cluster to obtain the loss-PGF of the cluster.¹⁵¹ These cluster PGFs are then again convoluted using FFT (and IFFT). The main idea behind this is that it is less demanding to compute the PGFs of the clusters because these are still relatively small units (or portfolios) and the computationally most intensive FFT and IFFT over a high-dimensional vector representing the whole portfolio is performed only once.

In our test, we took a completely non-homogeneous sample portfolio of 1000 exposures

¹⁵⁰an example might be a portfolio of a few very large and many small exposures. In such a case setting v too high significantly deteriorates the precision because the rounding errors of the small exposures are relatively high and frequent.

¹⁵¹note that the clusters are conditionally-independent.

and computed the probability coefficients of the unconditional overall portfolio loss PGF $G(z)$, which is the main output of the model. We applied the standard FFT method and compared it with the clustered FFT method. The time spent on the calculation of the portfolio loss distribution is shown in the Figure 33. It is easy to see that the clustered approach completely outperforms the standard FFT approach. If the amount of names in every cluster was set to 1 or 1 000 the speeds are the same as for the non-clustered versions. This is not a surprise as these marginal specifications of the clustered FFT are technically equal to the non-clustered version. We found the clustered FFT setup with ≈ 50 names in the cluster as the best choice.¹⁵² While the computation took about 80 seconds with the standard FFT, it took less than 3 seconds with the clustered FFT (50 names per cluster). A natural question is whether the clustered method can be iterated further in the sense that we would calculate the FFT of '(mini) clusters in clusters'. This is, however, beyond the scope of our analysis and we regard the presented clustered FFT method as sufficient.

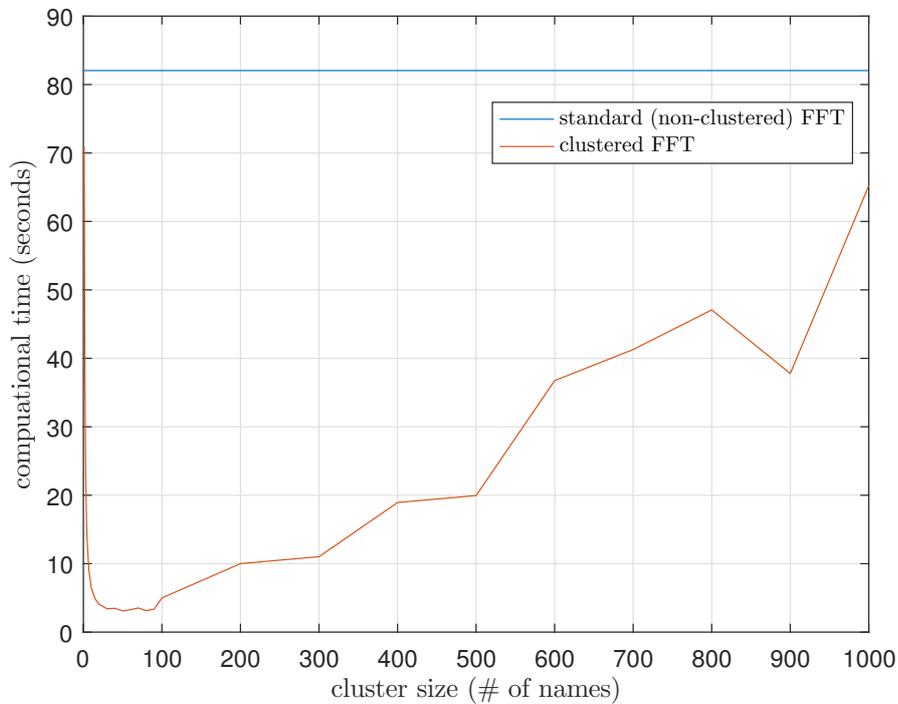


Figure 33: Runtime of the standard FFT vs clustered FFT on a portfolio of 1000 credits. Lower values indicate better performance.

For the environments where the FFT is *not* available such as Excel (VBA) or C++, we further propose a **recurrence relations approximation**, also taking into account the possible stochasticity of the RR/LGD. The idea is based on creating unique exposure bands of sizes E_m^* , $m = 1, \dots, M$ each of which is aggregating losses of exposures E_i of the size approximately E_m^* . Each band m has its own $X = x$ -conditional loss-probability

¹⁵²this result was still valid when we assumed $n > 1000$ credits in the portfolio.

generating function $G_m^*(z|x)$. The coefficient of z^n in $G_m^*(z|x)$, representing the probability of loss of n units in the band m , is denoted $L_n^m(x)$. The recurrence relations are summarized in the following Theorem.

Theorem IV.2.2 (Recurrence relation for loss probability in band m)

Let

$$c_m(\delta_j^*) = \left\lceil \frac{E_m^*(1 - \delta_j^*)}{v} \right\rceil, L_n^m(x) = \Pr[\text{loss in band } m = nv|x],$$

then

$$L_0^m(x) = \exp\left(-\sum_{i:E_i=E_m^*} Q_i(x)\right),$$

and for $n \geq 1$

$$L_n^m(x) = \frac{1}{n} \left(\sum_{i:E_i=E_m^*, j:0 < c_m(\delta_j^*) \leq n} \left(\sum_{i:E_i=E_m^*} Q_i(x) \Pr[\delta_i = \delta_j^*|x] \right) c_m(\delta_j^*) L_{n-c_m(\delta_j^*)}^m(x) \right).$$

To fully explain the main formula for $L_n^m(x)$ in the Theorem IV.2.2, let us break it down into several components. The first summation operator ranges over a selection of exposures E_i that belong to m th exposure band E_m^* and given those exposures it runs over all j (indexing recoveries δ_j^* for a given credit i) which guarantee loss given default less or equal to n units of v but strictly greater than 0.¹⁵³ The inner summation operator ranges again over the same i as the outer operator and calculates the cumulative probabilities that loss amounting to $c_m(\delta_j^*)$ occurs. This probability furthermore has two (conditionally-independent) components: the probability $Q_i(x)$ of i to default and the probability $\Pr[\delta_i = \delta_j^*|x]$ that the recovery hits δ_j^* . Put crudely, the inner operator says how many losses (expressed probability-wise) of the size $c_m(\delta_j^*)$ occur, while the outer operator iterates over all possible loss levels. The strength of the Theorem IV.2.2 is that it does not assume any particular specification of the factor functions driving the conditional default probabilities and possibly LGD/RR. LGD/RR still can be stochastic or simply a fixed constant.¹⁵⁴ A sample illustration shows in the Figure 34 that the approximation is almost accurate despite multiple defaults being allowed to occur for one name.

¹⁵³if $c_m(\delta_j^*) = 0$ is involved in the data (that is, the credit fully recovers after default), both equations in the Theorem IV.2.2 still holds but we must add one extra step at the very of the iteration procedure. This extra step involves setting

$$L_0^m(x) = \exp\left(-\sum_{i:E_i=E_m^*} Q_i(x)\right) + 1 - \exp\left(-\sum_{i:E_i=E_m^*, j:c_m(\delta_j^*)=0} Q_i(x) \Pr[\delta_i = \delta_j^*|x]\right),$$

after $L_n^m(x)$, $n \geq 0$ has already been calculated. The idea behind this is that the first $\exp(\cdot)$ still captures the probability of no default while the second term is the probability of defaults causing zero loss. Economically we are indifferent whether there is no default or there is a default with full recovery, hence we sum the probabilities of both these marginal cases. We however can not use this additional expression instead of $L_0^m(x)$ at the beginning of calculation of $L_n^m(x)$, $n \geq 0$ because we would distort the probabilities calculated recursively for $n > 0$.

¹⁵⁴in the case of constant LGD, $\Pr[\delta_i = \delta_j^*|x] = 1$ for δ_j^* equal to the constant LGD for the credit i .

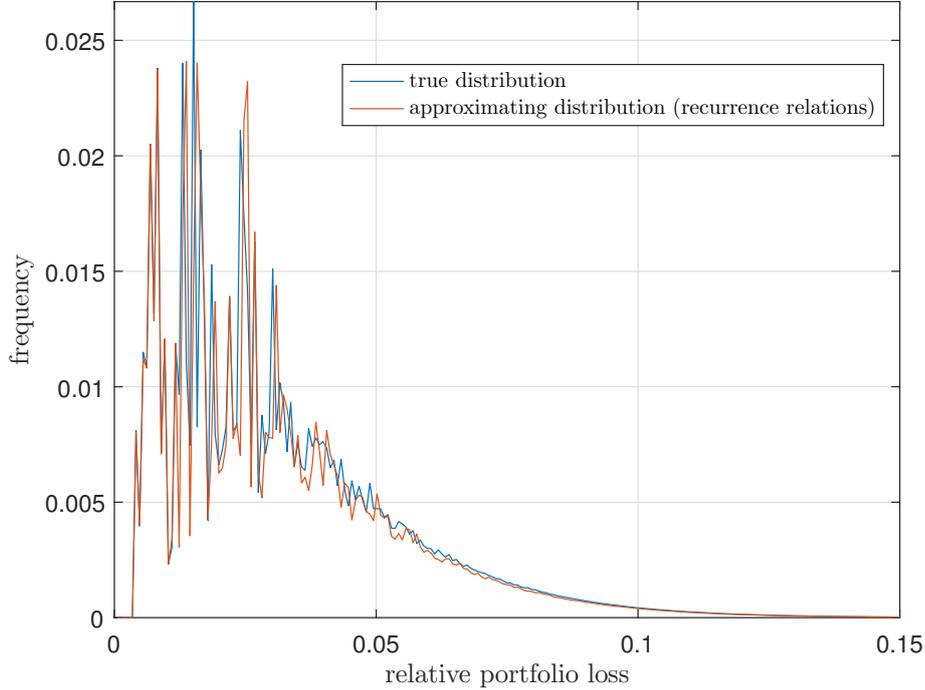


Figure 34: Comparison of the full (true) distribution and the approximating distribution through exposure bands. The number of credits in the portfolio is $n = 50$. The zero loss probability has been excluded from the chart.

IV.2.8 Dataset and benchmarking

For the purpose of a numerical conduct, we have generated a sample portfolio of 100 credits (loans/bonds) whose exposures range from 100 to 300. Further we have randomly generated PDs (q_i) in the range from 0% to 20% and the default correlations ρ_i ranging from 0 to 0.4. The recoveries δ_i have been sampled from Beta(2,5) distribution (and saved as fixed inputs) in the case of the fixed-RR model and in the case of the stochastic model we allow recoveries to be drawn from Beta(2,5) given randomly generated RR correlations ρ_i^δ in the range from 0 to ρ_i .¹⁵⁵ The choice of using Beta(2,5) can, for example, represent a recovery distribution of junior/subordinated bonds. In (Ou et al., 2016, Exhibit 8) the average long-term ultimate recovery rate is 0.2820 while our choice of Beta(2,5) gives expected recovery $2/(2 + 5) = 0.2857$. The distribution of PDs and RRs used for the numerical study is shown in the Figure 35.

Our main objective is to analyze the model outputs and make comparisons between the Gaussian and the double t copula. We in particular focus on EL, UL and ETL¹⁵⁶ quantities produced by the models. For completeness and we have also included the CreditRisk⁺ model (implemented according to the Technical Document (CreditSuisse, 1997)) as some

¹⁵⁵that is the recovery correlation ρ_i^δ is capped at the PD correlation ρ_i .

¹⁵⁶expected tail loss.

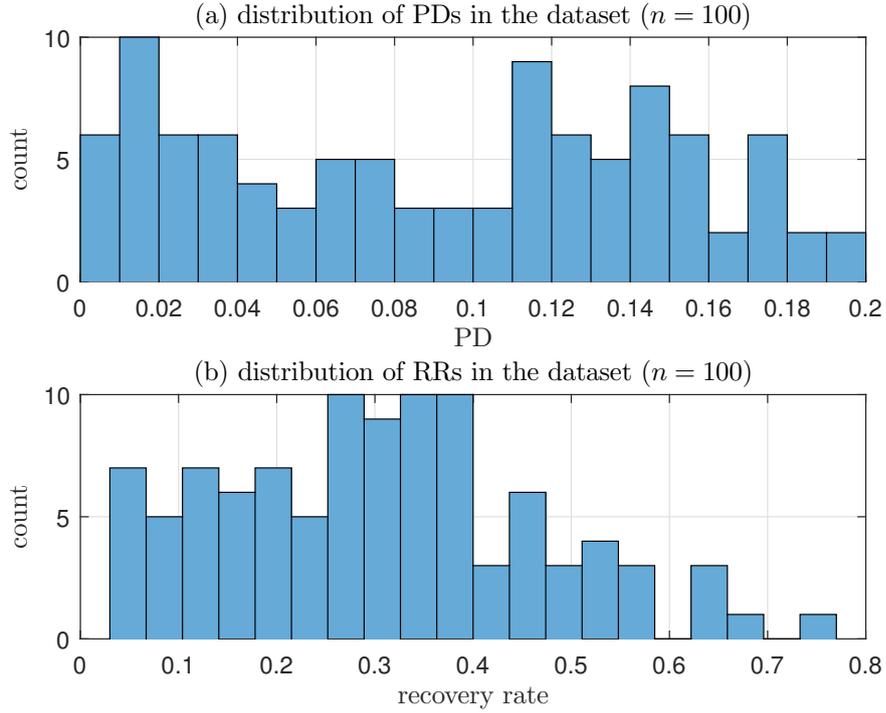


Figure 35: (a) histogram of PDs in the dataset, (b) histogram of recoveries used in the dataset.

objective and industry-wise applied benchmark. In CreditRisk^+ we actually use two setups: (i) low PD–volatility setup with PD volatilities of default rates σ_{PD} equal to $1/2$ of PDs,¹⁵⁷ and (ii) high PD–volatility setup with PD volatilities σ_{PD} equal to PD themselves. We also assume in CreditRisk^+ that all credits are grouped into one sector.¹⁵⁸

The most commonly used implementation in practice would probably be that of *fixed RR*. The full loss distributions for the Normal and double t setup¹⁵⁹ are shown in the Figure 34. While the EL in both models obviously coincides the UL of the double t implementation is twice the UL the Gaussian model. The ETL in the double t is also nearly a double of the ETL in the Gaussian model. This effect has been achieved by the fat-tailed double t distribution.

A more advanced approach to modeling is to assume that *recoveries are stochastic*. This is what we introduced in the Section IV.2.4. The idea is to use Beta(2, 5) distribution to model RR across all debtors instead of using fixed constants. Since recoveries, in general, can be correlated with a business cycle¹⁶⁰ it makes sense to assume that lower recoveries

¹⁵⁷this setting is used in the original 'demo' dataset accompanying the CreditRisk^+ technical document.

¹⁵⁸this also means that all credits are subject to one general factor variable X as in the proposed model framework.

¹⁵⁹ $\nu = 3$.

¹⁶⁰this is represented by the system variable X .

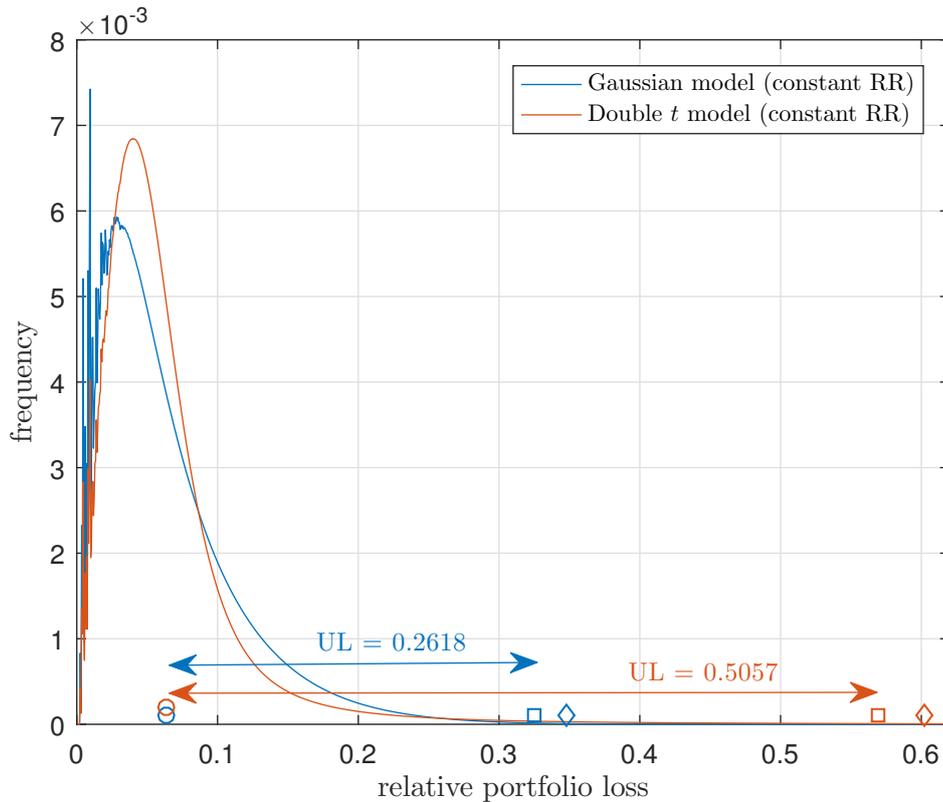


Figure 36: The loss distribution under the Gaussian (Normal) model and the double t with $\nu = 3$ model when RR is *constant*. \circ denotes EL, \square denotes $q_{99.9\%}$ loss and \diamond denotes ETL. Zero-loss probability has been excluded from the plot.

occur when the business cycle is in the downturn.¹⁶¹ The Figure 37 illustrates the Gaussian and double t loss distributions when the RR are allowed to be stochastic and only mildly correlated with X . The fact that the UL and ETL under the double t are roughly twice the UL and ETL under the Gaussian model still holds. What is, however, more interesting is to see the effect of the stochastic recoveries on the shift in UL and ETL compared to the fixed RR setup above. The absolute increment in these values is much more visible in the double t case. Therefore the double t approach is more reactive to the stochasticity of the RR than the Gaussian setup.

The overall results including the benchmark CreditRisk⁺ are summarized in the Table 20. There are, however, not many surprises. While the EL coincides across all the models in the test, the critical UL (and also ETL) figure differs significantly amongst the models. The double t model generally produces very high UL and ETL. Under the assumption that markets do not behave normally under stress, the double t model could represent a reasonable choice. On the other hand, capitalizing 50% – 65% of the portfolio notional is from the business perspective a very challenging objective to be met. CreditRisk⁺ shows

¹⁶¹for example, when there is market crash, the cash from the sale of assets might be worth less than when the market is booming due to low demand during the market crash (or recession, in general).

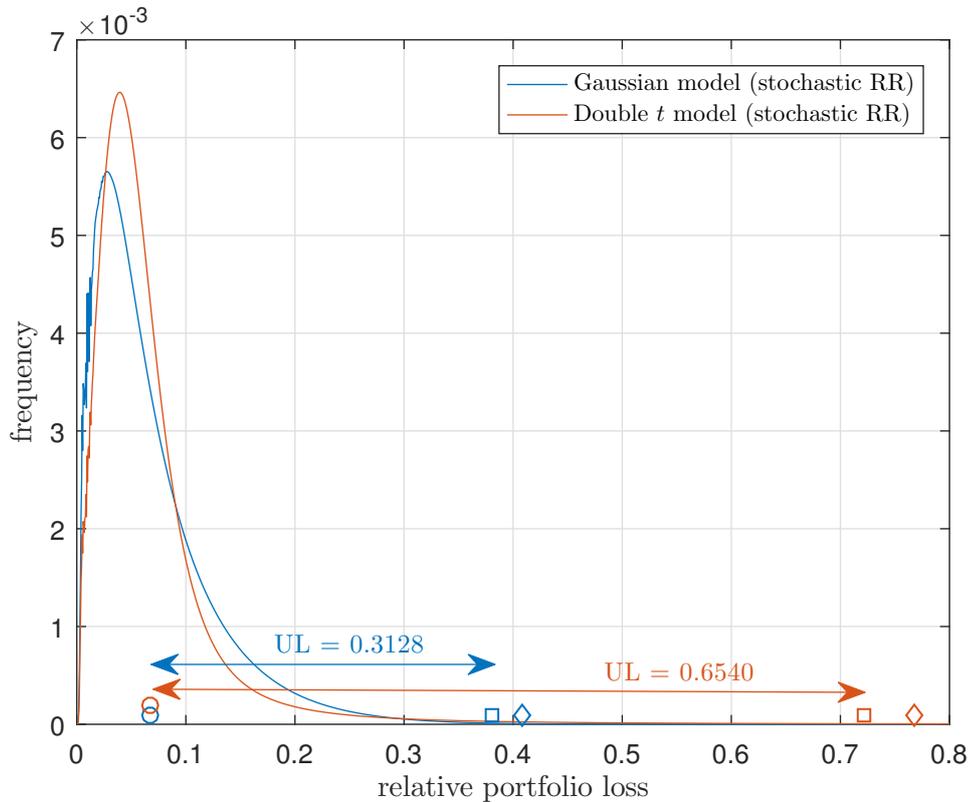


Figure 37: The loss distribution under the Gaussian (Normal) model and the double t with $\nu = 3$ model when RR is *stochastic*. \circ denotes EL, \square denotes $q_{99.9\%}$ loss and \diamond denotes ETL. Zero-loss probability has been excluded from the plot.

that UL is nearly a linear function of the volatility of the default rates σ_{PD} as UL nearly doubled when we doubled the σ_{PD} parameter. We were able to exceed the Gaussian-model UL when we allowed σ_{PD} to be very high, equal to PD itself. Having σ_{PD} very high can, however, also mean that the scoring/rating system is not reliable as observed defaults are too volatile compared to the predicted default rates. The ETL follows shortly after the $q_{99.9\%}$ loss in the Gaussian setup. This is because the loss distribution 'dies' quickly after $q_{99.9\%}$. In the double t approach the ETL is more distant from $q_{99.9\%}$. In CreditRisk⁺ ETL needs to be perceived very carefully because in CreditRisk⁺ repeated defaults can occur with a very low probability¹⁶² and thus the losses (and their probabilities) after $q_{99.9\%}$ can be also distorted by the 'repeated default' stigma. In contrast, in the proposed framework, we provide the *full*, true distribution and so repeated defaults are ruled out.¹⁶³

¹⁶²in turn the total portfolio loss can also exceed the total portfolio notional. This side effect penetrates the model through the approximation $\ln(1 + q_i(z - 1)) \approx q_i(z - 1)$ in (CreditSuisse, 1997, Equation (6)).

¹⁶³theoretically, the framework can generate losses greater than the notional. This is, however, related to the rounding procedure and not because of repeated defaults. Yet the approximating formula in the Theorem IV.2.2 allows for the repeated defaults with a marginal probability.

model	EL	UL	ETL	$q_{99.9\%}$ loss
Gaussian (constant RR)	0.0638	0.2628	0.3289	0.3266
Gaussian (stochastic RR)	0.0678	0.3128	0.4091	0.3806
double t (constant RR)	0.0638	0.5057	0.6025	0.5695
double t (stochastic RR)	0.0681	0.6540	0.7677	0.7221
CreditRisk ⁺ (low σ_{PD})	0.0615	0.1724	0.2548*	0.2339
CreditRisk ⁺ (high σ_{PD})	0.0615	0.3860	0.5127*	0.4475

Table 20: Summary table with loss statistics of various models in the study. The numerical figures are related to unit notional. (*) means ETL could be overvalued as repeated defaults can occur with a negligible probability.

IV.2.9 Concluding remarks

The project of this chapter was to present a semi-analytical model framework for portfolio credit risk calculation with adjustable features yet conceptually compatible with the general regulatory formulas. This could be practical for several reasons. (i) the semi-analyticality contributes to repeatability of exact results, often required by auditors or the regulator, (ii) the sensitivity analysis is simple,¹⁶⁴ (iii) the IRB methodology for the regulatory capital is conceptually close and hence there are not two possibly completely separated methods in a bank for the calculation of the regulatory and the economic capital for the same credit risk. Thus, it could be also more practical for discussions with the members of senior management who often prefer concepts and two different methods could be overly difficult to be communicated.

We have also shown that the methodology can be relatively painlessly adjusted to support environment with stochastic RR/LGD. The uncertainty of the RR/LGD is a big issue in practice. If a consumer loan has a high probability of attaining low RR and also a high probability of attaining high RR ('U-shape'), then the expected RR is about 50%, a value which would be hardly attained in the practice but would be tempting to be used if a constant has to be chosen. In this case the stochasticity of RR, linked to some distribution, is very useful and describes the best the real observations such that the bank can maximize the utility of the collected data.

As the intention was to provide a complete insight into the model structure, we also focused in detail on the model building blocks and the related technical details. We have discussed the aspects that affect the speed and also proposed approximating solutions for implementation in those environments that do not offer FFT functionalities.

Finally, it was shown how the double t model (which is attractive for the credit risk modeling) can be built-in. In the end we supported the methodology with a numerical experiment on a sample portfolio, also allowing for the benchmark against the established

¹⁶⁴frequently, senior managers ask the analytical department to check 'what happens if'. This is more difficult to do using simulations and could potentially lead to incorrect decisions.

CreditRisk⁺ model. The results we obtained were expectable and with no surprises. The stochasticity of the RR has brought an additional layer of uncertainty increasing UL in every setup. Finally, the double t generates a much higher UL than the Gaussian model, almost a double when $\nu = 3$.

A natural question is whether the figures produced by the double t model are applicable as they are much higher than the Gaussian model UL. This is a philosophical question and a discussion is necessary. First of all, in the stress conditions, the markets do not behave normally, hence the Gaussian setup might actually not be valid for the *crisis*. As already outlined above, the recent study of (Witzany, 2013) argues that the Gaussian setup significantly underestimates the risk and he comes to a conclusion that optimally the capital covering UL should be approximately twice as high. This is indeed also our finding (see the Table 20) when the double t model is used and in turn it could be a more appropriate model than the Gaussian copula for modeling credit portfolio risk. On the other hand, our testing has shown on the sample portfolio that to be prudent the bank would have to capitalize $1/2 - 2/3$ of the total credit exposure. This is a very conservative outcome. Also, one has to bear in mind that the confidence level 99.9% for the credit risk is very high. If a bank uses 99.9% confidence level then such a bank should default once in 1 000 years, yet a typical bank has PD higher than 0.1%. However, the level 99.9% for the credit risk might have been set to compensate the deficiencies of the Gaussian copula by forcing the Gaussian copula to generate higher capital requirements than what would correspond to bank's targeted rating (and PD).

Conclusion

The exposition provided in this thesis covers a broad area of the credit risk modeling with the emphasis on the pricing, where possible. The main motivation for this general coverage was to provide integrated, self-contained text on credit risk, not just focusing on a particular problem but make the presentation rather connected and context-based. Our main, general objective was to provide a sufficient insight into the credit risk modeling and pricing methodologies, extend them, develop own methods and support this with real-world applications. We also self-imposed a requirement on a balance between the theory and the applications in order to make the exposition complete while sufficiently focused on the applications.

In this text we presented several findings which were highlighted. In the Part II we summarized some standard results, built up a general pricing formula which supports a variety of applications. We also conducted a numerical study to answer the question whether the option market contains better credit information than the stock market. This has proved to be true, based on the data available.

The Part III was dedicated to the reduced-form credit models. We again revisited the theory and illustrated that many seemingly complicated problems can be efficiently solved using the main pricing equation that we outlined in a necessary detail. We also presented several more practical versions of this main pricing equation which might be more suitable for the actual implementation. We also provided several extensions, mainly the CVA equation and the Common Poisson Shock model generating higher credit correlations than the usual models. Finally, in the second section of the Part III, we have introduced a completely new two-level tree structure capable of pricing various, possibly path-dependent securities subject to default risk and we also numerically verified the outcomes using a combination of standard methods.

Portfolio modeling problems were discussed in the last Part IV. We firstly analyzed the CDO pricing problem, outlined the current market practice and its imperfections. This was accompanied by a suggestion about the numerical implementation, generating the results more efficiently than the conventional methods. Several CDO pricing methodologies were then discussed and we have also shown how to approach the pricing through the Archimedean copulas which less commonly appear in the CDO modeling and usually just stay in the theory. Our main contribution, however, was at the account of the double t copula which has proved to perform very well for the CDO pricing. The double t copula then appeared once again in the context of the portfolio risk modeling. We designed a simple, yet modular framework for modeling credit risk such that the general ideas ensure a high degree of compatibility with the regulatory formulas and the outcomes are reproducible. In this framework we also employed the double t copula and moreover presented its analytical solution for a special parameter case. The results generated with the double t copula seem to capture the behaviour of the credit market better than the normality-based models. We also derived an approximating distribution allowing to ease the computations on devices where the sophisticated functionalities are limited or not available.

A The fixed-income toolkit

In this section we review some important results from the fixed income. These results will accompany the rest of the thesis.

A.1 Basic continuous-time processes

Definition A.1 (The short-rate r)

The short interest rate $r(t)$ is defined as the interest rate observed at the time t prevailing over an infinitesimally short time-period $[t, t + dt]$, and so more precisely

$$r(t) = r(t, t + dt).$$

Definition A.2 (Instantaneous forward-rate f)

The instantaneous forward-rate contracted at time t and prevailing for the infinitesimal time period $[T, T + dt]$ is equal to

$$f(t, T) = -\frac{\partial \ln B(t, T)}{\partial T}, \quad (\text{A.1})$$

with the special property that

$$f(t, t) = r(t, t + dt) = r(t).$$

Definition A.3 (Money market account B)

A money market account B is a risk-free asset with unit initial balance ($B(0) = 1$) instantaneously growing at the rate r , that is

$$B(t) = B(0) \exp \left(\int_0^t r(u) du \right) = \exp \left(\int_0^t r(u) du \right).$$

Definition A.4 (Zero-coupon T bond $B(\cdot, T)$)

A zero-coupon default-free T -bond $B(\cdot, T)$ is an asset that pays 1 at T almost surely. The price of a zero-bond is connected to r and f by the following two fundamental equations

$$B(t, T) = B(t) \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{B(T)} \middle| \mathcal{F}(t) \right] = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^T r(u) du \right) \middle| \mathcal{F}(t) \right], \quad (\text{A.2})$$

and

$$B(t, T) = \exp \left(- \int_t^T f(t, u) du \right).$$

In many applications, and particularly in the HJM framework, we need to derive other processes given the dynamics of the the instantaneous forward rate f is known. In these cases, the following two propostions are very useful.

Proposition A.1 (Forward-rate dynamics implies bond dynamics)

If the forward-rate $f(t, T)$ follows

$$df(t, T) = \alpha(t, T)dt + \sigma(t, T)dW(t).$$

then the bond price $B(t, T)$ follows

$$dB(t, T) = B(t, T)m(t, T)dt + B(t, T)v(t, T)dW(t),$$

with

$$\begin{aligned} m(t, T) &= r(t) - \int_t^T \alpha(t, u)du + \frac{1}{2}|S(t, T)|^2 \\ v(t, T) &= S(t, T), \end{aligned}$$

in which furthermore

$$S(t, T) = - \int_t^T \sigma(t, u)du.$$

Proposition A.2 (Forward-rate dynamics implies short-rate dynamics)

If the forward-rate $f(t, T)$ follows

$$df(t, T) = \alpha(t, T)dt + \sigma(t, T)dW(t),$$

then the short-rate r follows

$$dr(t) = (f_T(t, t) + \alpha(t, t)) dt + \sigma(t, t)dW(t).$$

A.2 Fundamentals of the HJM framework

In the above short section dedicated to the basic fixed-income market processes we assumed that the basic component of an interest-rate model is the instantaneous rate r . This, however, might not be practical for two main reasons:

- It is difficult to justify that the money-market account B is dominated by just one variable r .
- The term-structure of the forward rates ($T \in [t, T^*] \mapsto f(t, T)$) obtained by first specifying the short-rate is sometimes very unrealistic. If one then wants to obtain a realistic term-structure of the forward rates, this might require a complicated specification of r .

These limitations of the 'short-rate' r modeling can be overcome in the Heath-Jarrow-Morton (HJM) framework (Heath et al., 1990, 1991, 1992), where the basic component is the forward-rate $f(t, T)$ instead of r . The main benefit of this approach is that by specifying the volatility function of the forward rate one gets a complete, arbitrage-free dynamics of the fixed-income instruments (applying the general Propositions A.1 and A.2).

Let us start with the general SDE describing the dynamics of the forward-rate $f(t, T)$ directly under \mathbb{Q} . $f(t, T)$ follows

$$df(t, T) = \alpha(t, T)dt + \sigma(t, T)dW^{\mathbb{Q}}(t), \tag{A.3}$$

where σ is allowed to be a row vector and $W^{\mathbb{Q}}$ is allowed to be a column vector. The main question here is under what condition is the model arbitrage-free. Answer to this question is addressed by the *HJM drift restriction*.

Lemma A.1 (HJM drift restriction)

Under \mathbb{Q} the drift α of the forward rate $f(t, T)$ must necessarily satisfy

$$\alpha(t, T) = \sigma(t, T) \int_t^T \sigma(t, u)^T du.$$

The Lemma A.1 introducing the HJM drift restriction is the main result of the HJM approach. It specifies, what \mathbb{Q} -dynamics is followed by $f(t, T)$ having specified the volatility term $\sigma(t, T)$.

To build a HJM model with $f(t, T), r(t), B(t, T)$ one typically proceeds as follows:

1. Select a volatility function $\sigma(t, T)$ for the forward-rate *of your choice*.
2. Apply the HJM drift restriction (Lemma A.1) to compute the drift $\alpha(t, T)$ of $f(t, T)$.
3. Integrate the forward-rate dynamics (A.3) from 0 to t to obtain $f(t, T)$, i.e. compute

$$f(t, T) = f(0, T) + \int_0^t \alpha(u, T) du + \int_0^t \sigma(u, T) dW^{\mathbb{Q}}(u).$$

4. Find the bond price \mathbb{Q} -dynamics by applying the Proposition A.1. Notice here that $m(t, T) = r(t)$ by the definition of \mathbb{Q} .¹⁶⁵
5. Find the short-rate r \mathbb{Q} -dynamics by computing $f_T(t, T)$ and then apply the Proposition A.2.

A.3 Caps, floors, swaptions and Black formulas

In various applications we are in the need to calibrate an interest-rate model to the market quotes. Since a calibration to the bond prices is often not enough (for example in the case when the interest-rate model provides an automatic perfect fit to the yield curve), an additional source of market information is necessary. Due to a high liquidity, we work with the information embedded in the interest-rate derivatives, namely caps/floors and swaptions as there is a huge and liquid market for these instruments. Although there are various approaches to pricing of these derivatives, the market standard is to use market quotes in the form of implied volatilities and use them in some version of Black's formula to recover the price of the derivative.

In what follows let us define a *simply compounded* LIBOR rate $L(t, T) = L(t, T, T + \delta)$ which is a rate that resets at T and pays at $T + \delta$ using the formula

$$L(t, T) = \frac{1}{\delta} \left(\frac{B(t, T)}{B(t, T + \delta)} - 1 \right).$$

We note that for the case when $t \geq T$, $L(t, T)$ is a fixed number corresponding to $L(T, T)$, that is it corresponds to the spot LIBOR observed at the reset date T . **Interest-rate**

¹⁶⁵note that according to Proposition A.1 for $dB(t, T)$ only the term $\sigma(t, T)$ from the forward-rate dynamics is needed. Anything else is not needed so the two prior steps can be skipped.

cap is an instrument that consists of simpler instruments called *caplets* c_i . Having quoted $xM \times yY$ cap means that the cap runs for y years and pays off every x months the excess of x -month LIBOR over some rate (cap) R . The x months are represented here in years by the tenor δ . We thus have a set of calendar times $T_i = i\delta, i = 1, \dots, n$ with $t = 0$ being the current time and T_n being the cap termination date. T_1 is the first reset date and T_{n-1} is the last reset date. In total there are always $n - 1$ payoffs. Thus, for a $3M \times 5Y$ cap, $n = 5/0.25 = 20$ and there are 19 payoffs (caplets $i = 1, \dots, n - 1$).

Once the LIBOR $L(T_i, T_i)$ has been fixed at T_i , the i th caplet is *supposed to pay* at $T_{i+1} = T_i + \delta$ the amount

$$\delta \max[L(T_i, T_i) - R, 0].$$

We say that the caplet i *covers* the period $[T_i, T_i + \delta]$. This payoff is economically equivalent to a discounted payoff at the time of the reset T_i equalling to

$$c_i(T_i) = B(T_i, T_i + \delta) \delta \max[L(T_i, T_i) - R, 0], i = 1, \dots, n - 1. \quad (\text{A.4})$$

The present value of the caplet c_i under \mathbb{Q} equals

$$c_i(t) = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{T_i} r(u) du \right) c_i(T_i) \middle| \mathcal{F}(t) \right],$$

and so the cap value $C(t)$ equals

$$C(t) = \sum_{i=1}^{n-1} c_i(t).$$

For the **interest-rate floors**, this is a pure analogy and we have that a floor F consists of $n - 1$ floorlets f_i with payoffs $f_i(T_i), i = 1, \dots, n - 1$ given as

$$f_i(T_i) = B(T_i, T_i + \delta) \delta \max[R - L(T_i, T_i), 0],$$

The present value of the floorlet f_i equals

$$f_i(t) = \mathbb{E}^{\mathbb{Q}} \left[\exp \left(- \int_t^{T_i} r(u) du \right) f_i(T_i) \middle| \mathcal{F}(t) \right],$$

and the present value of floor F is

$$F(t) = \sum_{i=1}^{n-1} f_i(t).$$

It is customary amongs the market practitioners to use **Black 76'** formula to price caps/caplets and floors/floorlets. In Black 76, the price of caplet c_i (floorlet f_i) is given as

$$\begin{aligned} c_i(t) &= \delta B(t, T_i + \delta) [L(t, T_i)N(d_1) - RN(d_2)] \\ f_i(t) &= \delta B(t, T_i + \delta) [RN(-d_2) - L(t, T_i)N(-d_1)] \\ d_1 &= \frac{\ln \left(\frac{L(t, T_i)}{R} \right) + \frac{1}{2} \sigma_i^2 (T_i - t)}{\sigma_i \sqrt{T_i - t}} \\ d_2 &= d_1 - \sigma_i \sqrt{T_i - t}, \end{aligned}$$

where σ_i is the implied (or Black) volatility for i th caplet (floorlet). Typically, the individual caplets (floorlets) in a cap (floor) have different σ_i . This would be, however, impractical for the quotes. The market practice is to quote one implied volatility¹⁶⁶ $\sigma_i \equiv \sigma$ which is used for pricing of all caplets (floorlets) in a given cap (floor).¹⁶⁷

The second widely traded interest-rate derivative is a **swaption**. Swaption $S_w(t, K)$ is an option to enter IRS whose fixed leg pays coupons at the annual rate K . The swaption is characterized by two critical dates. T_0 is a date when the swaption terminates (and the underlying IRS starts) while T_n is a date when the underlying IRS matures. To keep these information together, we often write ' $T_0 \times (T_n - T_0)$ swaption' for the swaption expiring at T_0 on IRS starting at T_0 and maturing at T_n . The component $(T_n - T_0)$ has a customary name 'IRS tenor'.

Specifically, the payer swaption ($w = 1$) is an option to become a fixed rate payer (at rate K) in the underlying IRS, while the receiver swaption ($w = -1$) gives its holder the right to enter to an IRS as a receiver of the fixed rate (K). Let us denote $\text{IRS}_w(T_0, K)$ the value of the IRS at its inception date, when the fixed leg is assumed to pay at rate K . If $w = 1$ then such value represents the value from the perspective of fixed receiver, while $w = -1$ means the value as seen from the fixed payer.

The payoff of a swaption at time its exercise date T_0 equals

$$S_w(T_0, K) = \max[\text{IRS}(T_0, K, w), 0], \quad (\text{A.5})$$

where $w = -1$ is payer swaption and $w = 1$ is a receiver swaption. To approach the pricing of swaptions, we need to define the price of the IRS first. We note that given we have specified the IRS coupon rate K , the value of such IRS prior and at the first reset date T_0 equals

$$\text{IRS}_w(t, K) = \begin{cases} K\delta \sum_{i=1}^n B(t, T_i) - B(t, T_0) + B(t, T_n), & \text{if } w = 1 \text{ (fix receiver)} \\ B(t, T_0) - B(t, T_n) - K\delta \sum_{i=1}^n B(t, T_i), & \text{if } w = -1 \text{ (fix payer)} \end{cases}, t \leq T_0, \quad (\text{A.6})$$

regardless of the tenor of the floating LIBOR rate being reset in the IRS, yet taking into account the length of the fixed coupon periods, expressed as a equidistant year fraction δ . For pricing of S_w in terms of the IRS, it would be overly difficult to specify the dynamics of all the underlying rates that drive the bond prices $B(\cdot, T_i)$ which determine the value of the IRS in question. For this reason the market has adopted the convention to price the swaptions by comparing the 'underlying IRS' with a 'hypothetical IRS' that would pay the fixed coupons at rate $R(t)$ which is defined as

$$R(t) = \frac{B(t, T_0) - B(t, T_n)}{\delta \sum_{i=1}^n B(t, T_i)},$$

and this rate $R(t)$ is referred to as the *forward swap rate*. R is such a value of the fix

¹⁶⁶ σ_i is usually named *spot* volatility while σ is referred to as *flat* volatility.

¹⁶⁷packing all σ_i into one flat volatility σ to value a cap (floor) is in fact similar to pricing a bond using its yield to maturity rather than indivial rates for every cash-flow.

rate K for which the IRS is zero valued at time $t \leq T_0$. The main trick in connecting the swaption's payoff (A.5) with the forward swap rate $R(t)$ is based on the the following idea

$$\begin{aligned} S_w(T_0, K) &= \max[\text{IRS}_w(T_0, K), 0] \\ &= \max[\text{IRS}_w(T_0, K) - 0, 0] \\ &= \max[\text{IRS}_w(T_0, K) - \text{IRS}_w(T_0, R(T_0)), 0], \end{aligned} \quad (\text{A.7})$$

where we used the property that $\text{IRS}_w(T_0, R(T_0)) \equiv 0$. Substituting into (A.7) for the IRS terms from (A.6) and using the definition of the rate R gives the payoff of S_w in terms of R as

$$S_w(T_0, K) = \begin{cases} \max[K - R(T_0), 0] \delta \sum_{i=1}^n B(t, T_i), & \text{if } w = 1 \text{ (receiver swaption)} \\ \max[R(T_0) - K, 0] \delta \sum_{i=1}^n B(t, T_i), & \text{if } w = -1 \text{ (payer swaption)} \end{cases}.$$

Now, there is only *one* stochastic variable $R(T_0)$ in place. Similarly to caps and floors, the market practice is to use Black's formula which assumes $R(T_0)$ is a lognormally distributed variable and to use the quotes of the implied volatility to retrieve the value of $S_w(t, K)$. **Black's formula for swaptions** reads

$$S_w(t, K) = \begin{cases} (KN(-d_2) - R(t)N(-d_1)) \delta \sum_{i=1}^n B(t, T_i), & \text{if } w = 1 \text{ (receiver swaption)} \\ (R(t)N(d_1) - KN(d_2)) \delta \sum_{i=1}^n B(t, T_i), & \text{if } w = -1 \text{ (payer swaption)} \end{cases},$$

where

$$\begin{aligned} d_1 &= \frac{\ln\left(\frac{R(t)}{K}\right) + \frac{1}{2}\sigma^2(T_0 - t)}{\sigma\sqrt{T_0 - t}} \\ d_2 &= d_1 - \sigma\sqrt{T_0 - t}. \end{aligned}$$

As usual in the Black-like formulas, σ is the implied (Black) volatility for the given swaption.

A.4 Caps, floors and swaptions in the short-rate models

As already outlined in the Section (A.3), in our applications we want to use the quotes of the interest-rate derivatives as a reliable source of information for the calibration of the interest-rate models. In this section we provide connections between these instruments and the interest-rate models. We will specialize on the most common Gaussian case, generally specified as

$$dr(t) = \kappa(t)(\theta(t) - r(t))dt + \sigma(t)dW^{\mathbb{Q}}(t),$$

which provides a perfect fit to the observed yield curve when

$$\theta(t) = \frac{1}{\kappa(t)} \frac{\partial f(0, t)}{\partial t} + f(0, t) + \frac{1}{\kappa(t)} \int_0^t \sigma^2(u) e^{-2 \int_u^t \kappa(s) ds} du.$$

Since, as it turns out later, we represent all the derivatives pricing problems as problems involving just the pricing of the pure discount bonds, we necessarily need a formula for the bond prices. In many practical applications, however, it is more appropriate to work with $x(t) = r(t) - f(0, t)$ rather than with $r(t)$ itself because application of this simple transformation eliminates the f -partial.¹⁶⁸

Theorem A.1 (Dynamics and bond prices in terms of $x(t) = r(t) - f(0, t)$)

Let

$$x(t) = r(t) - f(0, t).$$

Then x follows

$$dx(t) = (y(t) - \kappa(t)x(t))dt + \sigma(t)dW^{\mathbb{Q}}(t), x(0) = 0,$$

The bond prices can be recovered by the formula

$$B(t, T) = \frac{B(0, T)}{B(0, t)} \exp \left(-x(t)G(t, T) - \frac{1}{2}y(t)G(t, T)^2 \right).$$

The dynamics of $B(t, T)$ is stated as

$$\frac{dB(t, T)}{B(t, T)} = r(t)dt + v(t, T)dW^{\mathbb{Q}}(t), v(t, T) = -\sigma(t)G(t, T).$$

The terms y and G are given as

$$y(t) = \int_0^t \sigma^2(u) e^{-2 \int_u^t \kappa(s) ds} du$$

$$G(t, T) = \int_t^T e^{-\int_t^u \kappa(s) ds} du.$$

Finally, we can define bond option formulas for the above introduced Gaussian model.

Theorem A.2 (Bond option)

Let $t < T < s$, where T denotes the time of expiration of option on a s -maturity zero bond. Then the call and the put options on that s -bond have the pricing formulas

$$C(t, T) = B(t, s)N(d_1) - KB(t, T)N(d_2)$$

$$P(t, T) = KB(t, T)N(-d_2) - B(t, s)N(-d_1),$$

where

$$d_1 = \frac{\ln \left(\frac{B(t, s)}{KB(t, T)} \right) + \frac{1}{2}v}{\sqrt{v}}, d_2 = d_1 - \sqrt{v}$$

$$v = \int_t^T (v(u, s) - v(u, T))^2 du.$$

¹⁶⁸usually, $f(0, t)$ is not a perfectly smooth curve and thus it might be complicated to work with its derivative.

Having outlined the bond options, it is easy to connect them with caplets (floorlets). It can be shown that the c_i 's payoff (A.4) (and analogously also the floorlet f_i 's payoff) which covers the period $[T_i, T_i + \delta]$ can be written as

$$\begin{aligned} c_i(T_i) &= (1 + \delta R) \max \left[\frac{1}{1 + \delta R} - B(T_i, T_i + \delta), 0 \right] \\ f_i(T_i) &= (1 + \delta R) \max \left[B(T_i, T_i + \delta) - \frac{1}{1 + \delta R}, 0 \right]. \end{aligned}$$

From these, it is easy to see that a caplet (floorlet) is effectively a put (call) bond option with a scaled 'notional' and a modified strike. We present these formulas explicitly in the following theorem.

Theorem A.3 (Caplets and floorlets)

Caplet c_i (floorlet f_i) covering the period $[T_i, T_i + \delta]$ at rate R has the value

$$\begin{aligned} c_i(t) &= B(t, T_i)N(-d_2) - (1 + \delta R)B(t, T_i + \delta)N(-d_1) \\ f_i(t) &= (1 + \delta R)B(t, T_i + \delta)N(d_1) - B(t, T_i)N(d_2), \end{aligned}$$

where

$$\begin{aligned} d_1 &= \frac{\ln \left(\frac{(1 + \delta R)B(t, T_i + \delta)}{B(t, T_i)} \right) + \frac{1}{2}v}{\sqrt{v}}, \quad d_2 = d_1 - \sqrt{v} \\ v &= \int_t^{T_i} (v(u, T_i + \delta) - v(u, T_i))^2 du. \end{aligned}$$

We are left with the definition of swaption pricing using interest-rate models. This is the subject of the last part of this section.

The case of **swaptions** is, however, much more complicated compared to caplets (floorlets). To see this, let us focus on the swaption's payoff at T_0 . Combining (A.5) with (A.6) and setting $t = T_0$ gives

$$S_w(T_0, K) = \begin{cases} \max \left[K\delta \sum_{i=1}^n B(T_0, T_i) + B(T_0, T_n) - 1, 0 \right], & \text{if } w = 1 \text{ (receiver swaption)} \\ \max \left[1 - B(T_0, T_n) - K\delta \sum_{i=1}^n B(T_0, T_i), 0 \right], & \text{if } w = -1 \text{ (payer swaption)} \end{cases}.$$

Obviously, the random variables $B(T_0, \cdot)$ in the $\max[\cdot]$ function can not be factored out and so the swaption can not be trivially decomposed into the individual bond options as it was possible for the caps and floors. Apparently, however, there exists some critical level of the interest-rate $r^*(T_0) = r^*$ for which the swaption is ATM. As it is more convenient to work with $x(t) = r(t) - f(0, t)$ rather than $r(t)$, we analogously say that there exists a corresponding critical value $x^* = r^* - f(0, T_0)$. Let us write $B(T_0, T_i) = B(T_0, T_i, x(T_0))$ for a while. Then the critical level x^* can be found by a one-dimensional numerical search over various $x(T_0)$ for which $S_w(T_0, K) = 0$, i.e.

$$1 = B(T_0, T_n, x^*) - K\delta \sum_{i=1}^n B(T_0, T_i, x^*).$$

Applying the Jamshidian's trick, we must construct hypothetical strikes equal to

$$K_i = B(T_0, T_n, x^*), i = 1, \dots, n.$$

The main idea of the trick is to decompose the option on a portfolio of bonds¹⁶⁹ into a portfolio of options on zero bonds with the created strikes K_i . It eventually turns out that

$$S_w(T_0, K) = \begin{cases} \max[B(T_0, T_n) - K_n, 0] + K\delta \sum_{i=1}^n \max[B(T_0, T_i) - K_i, 0], & \text{if } w = 1 \\ \max[K_n - B(T_0, T_n), 0] + K\delta \sum_{i=1}^n \max[K_i - B(T_0, T_i), 0], & \text{if } w = -1 \end{cases}. \quad (\text{A.8})$$

Now, it is easy to see that the value of swaption $S_w(t, K)$ is equal to portfolio of call (put) options with strikes $K_i, i = 1, \dots, n$ on zero bonds maturing at T_i in the case of receiver (payer) swaption. To evaluate (A.8) one can thus use the bond option formulas presented in the Theorem A.2.

B Reduced Convolution

Arguably the fastest way of computing the PGF $G(z|x)$ of the whole portfolio in the Section IV.2.3 is through the FFT method. This, however, is not always possible if the FFT functionalities are not available. In such a case, one can use either the approximating formula (see the Theorem IV.2.2) and then convolute the loss-PGFs of the 'bands' $G_m^*(z|x)$ to get $G(z|x)$. When the approximation is not applied, all the individual loss-PGFs $G_i(z|x), i = 1, \dots, n$ have to be convoluted. Hence, if one assumes away the FFT method, the convolution can not be avoided. Since the loss-PGFs G_i we are working with are specific - having many zero elements, we propose an optimized convolution algorithm to leverage the benefits of this special property. We call this method the *reduced convolution*, because of the fact that it reduces the standard convolution only to the elements which are non-zero.

Without the loss of generality, for simplicity assume that there are only two PGFs, G_A, G_B and the overall 'portfolio' PGF is G_C . We assume that G_A is a 'complicated' PGF with many non-zero coefficients, and G_B is a 'simple' PGF, with almost all coefficients equal to zero.¹⁷⁰ Let $\mathbf{A}, \mathbf{B}, \mathbf{C}$ denote the vector of coefficients in G_A, G_B, G_C , respectively. The dimensions of these vectors are d_A, d_B and $d_C = d_A + d_B - 1$. Then

$$G_C(z|x) = G_A(z|x) G_B(z|x),$$

has the coefficients in $\mathbf{C} = \mathbf{A} * \mathbf{B}$ given as

$$C_k = \sum_{j=0}^k A_j B_{k-j}, k = 0, 1, \dots, d_C - 1.$$

¹⁶⁹one can think of the 'portfolio of bonds' as a coupon-bearing bond.

¹⁷⁰in practice G_A is PGF of previously convoluted G_i s, while G_B is a loss-PGF of a single credit to be convoluted with G_X .

This is the general formula for the convolution '*'. The problem with this simple formula is that to fill \mathbf{C} , we need all $C_k, k = 0, 1, \dots$ and this results into a quadratic loop. The idea behind the reduced convolution is to organize the convolution such that the elements in \mathbf{B} are not called repeatedly but each element only once. We propose to formulate the convolution in the following way:

$$\mathbf{C} = \mathbf{A} * \mathbf{B} = \sum_{k=0}^{d_C-1} \mathbf{V}_k, \mathbf{V}_k = (0, \dots, 0, B_k A_0, B_k A_1, \dots, B_k A_{d_A-1}, 0, \dots, 0)^T, \quad (\text{B.1})$$

where each vector \mathbf{V}_k has the dimension $d_A + d_B - 1$. The first non-zero element of \mathbf{V}_k starts at the index k and the last non-zero element is at index $k + d_A - 1$. Notice that each vector \mathbf{V}_k can be technically formed very quickly as it only consist of a vector $(B_k A_0, B_k A_1, \dots, B_k A_{d_A-1})^T = B_k \mathbf{A}$ placed in a particular position, surrounded by zeros. Since in (B.1) we have arranged the convolution as a single sum of vectors, each depending on $B_k \mathbf{A}$ in the sum we can disregard all the vectors which have all elements zeros. Given at least one non-zero element in \mathbf{A} , we will have zeros-vector \mathbf{V}_k if and only if $B_k = 0$. Since zero-vectors have no contribution to the sum, we can drop these from the summation (B.1) and we have

$$\mathbf{C} = \mathbf{A} * \mathbf{B} = \sum_{k \geq 0, k: B_k \neq 0} \mathbf{V}_k, \mathbf{V}_k = (0, \dots, 0, B_k A_0, B_k A_1, \dots, B_k A_{d_A-1}, 0, \dots, 0)^T,$$

that is we only sum vectors indexed k , when $B_k \neq 0$. Because \mathbf{B} by our assumption holds only a few (typically a two for fixed RR model) non-zero B_k , the convolution is very simple to be computed. The Box B provides a code for this 'reduced convolution' formula.

Box B.1: MATLAB code for the *reduced convolution*

```
function Z = reduced_conv(X, Y)
% reduced convolution algorithm for loss-PGF
% returns Z = conv(X, Y), where X and Y are loss-PGFs

%% inputs
% X = the 'long' vector including less zero elements
% Y = the 'short' vector including many zero elements

%% calculation
% identify where in Y (except for the first element)
% are non-zero elements (probabilities)
dimX = numel(X); non_zeros_idx = find(Y(2:end)>0)+1;
num_non_zeros_Y = numel(non_zeros_idx);
% create the output loss-PGF vector Z; set the initial values in Z to X*(1-Q)
Z = zeros(dimX+numel(Y)-1, 1); Z(1:dimX,1) = X*Y(1);
% iterate over all other non-zero elements in Y
% and add the probabilities X*Y(i) to Z
for i = 1:num_non_zeros_Y
    idx_range = non_zeros_idx(i):non_zeros_idx(i)+dimX-1;
    Z(idx_range,1) = Z(idx_range,1) + X*Y(non_zeros_idx(i));
end

end
```

When we benchmarked this algorithm in MATLAB on \mathbf{A} being 5 000 000-dimensional vector, and \mathbf{B} being 3 000 000-dimensional vector (with three non-zero entries), we measured the runtime 0.2s with reduced convolution, 0.45s with the standard MATLAB convolution¹⁷¹ and runtime 3.2s with the FFT method. We have thus reduced the runtime to a half of the MATLAB's inbuilt algorithm. Notice that FFT is slower here but on the other hand it allows for multiple vectors being convoluted at one time which eventually outperforms any other method when a sufficient amount of vectors is convoluted.¹⁷²

C Numerical solutions to PDEs

A substantial part of the Section III.2.1 was based on numerical solutions to PDEs. In this part of the Appendix we reveal in detail how the PDEs were solved through own numerical schemes. There were two main issues we encountered while working out the numerical schemes for the PDEs. Firstly, the classical Dirichlet and Neumann boundary conditions (BCs), typically referred to by the literature (see e.g. (Wilmott et al., 1995, Section 3.7), (Pikulin, 2001), (Duffy, 2006, Section 1.2.1)) could not be used, in general. This is because in the field of the reduced-form modeling it is 'nearly impossible' to determine *how close* is the credit from the default.¹⁷³ In turn, we are not able to specify

¹⁷¹0.45s was achieved in MATLAB 2016. In MATLAB 2012, we measured 24s.

¹⁷²also notice we can employ the clustered FFT method described in the Section IV.2.7.

¹⁷³this is actually a technical consequence of the *inaccessibility* of the default time τ^* . Recall the Definition III.1.2.

the value of the claim V when the intensity γ is high or low (at the upper or the lower boundary of γ). The second difficulty we had to deal with was the stability/convergence of the scheme.¹⁷⁴ The main objective of the scheme is *convergence*. A scheme, by virtue of the Lax equivalence theorem,¹⁷⁵ is *convergent if and only if it is consistent and stable at the same time*. The consistence¹⁷⁶ is obvious and we were able to maintain it without any difficulties but the same can not be said about the stability. This is also related to the BC issue. When BC can not be accurately specified a 'noise' enters the scheme and as these disturbances add up, the scheme might fail completely. We have, however, managed to tackle both these complications with significant adjustments in the scheme. Everything what follows is already customized so that these 'ghost' effects are prevented. Therefore, the specification of the schemes might differ from the conventional literature, typically dealing with some sort of optimal problems such as Black-Scholes which is a simple numerical problem in the finite difference scheme.

The PDE finite difference schemes that we recommend to use and are described below are based on the following three critical paradigms:

- (i) the setup of the BCs is done *implicitly* in such a way that neither Dirichlet nor Neumann BCs are used. The whole PDE is used as the BCs instead,
- (ii) all the finite differences are directed to face *inwards* into the finite difference scheme,
- (iii) the scheme is organized such that the spatial discretization error is $O(h^2)$.

To justify these paradigms, let us allow to provide a short comment. The points (i) and (ii) are related. The aim is to use a scheme that does not depend on the explicit specifications of the boundary conditions. There are two main reasons why this is desirable. Firstly, the same scheme¹⁷⁷ can be used for any PDE. For example, we can use the same 2D-PDE scheme for a defaultable bond and for a stock option (Heston model) *without any additional concern about the BCs*. Secondly, the explicit definition of BCs is sometimes (and in the field of credit very often) impossible. To overcome this, we assume that the best possible BC is the PDE itself. This, however, enforces some adjustments in the scheme, which in this case involves to ensure an appropriate 'directing' of the finite differences (point (ii)). The paradigm (iii) is a technical, stability supporting requirement which turned out to be sufficient for the pricing cases we dealt with. We found that the schemes were critically sensitive to the degree of accuracy. In turn it was necessary to provide a full specification of the derivatives everywhere on the grid, keeping the focus on the $O(h^2)$ accuracy.

C.1 One-dimensional PDEs

A typical general parabolic one-dimensional PDE looks like

$$\frac{\partial V}{\partial t} + \mu(t, x) \frac{\partial V}{\partial x} + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 V}{\partial x^2} + h(t, x) - r(t, x)V = 0,$$

¹⁷⁴these are two different but interrelated issues.

¹⁷⁵this is discussed in a footnote in the Section III.2.2.1.

¹⁷⁶this, heuristically, means that the scheme approximates the PDE solution.

¹⁷⁷obviously, with a different set of coefficients $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$.

with a certain behaviour at $t = T$ given by $V(T, x) = g(x)$. Let \mathcal{L} be an operator¹⁷⁸

$$\mathcal{L} = \mu(t, x) \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2}{\partial x^2} - r(t, x),$$

such that

$$\frac{\partial V}{\partial t} + \mathcal{L}V + h(t, x) = 0.$$

Let $\widehat{\mathcal{L}}$ be a discrete version of \mathcal{L} defined as

$$\mathcal{L}V = \widehat{\mathcal{L}}V + \varepsilon,$$

where ε is a discretization error of some order. In order to define operations through $\widehat{\mathcal{L}}$, the time space and the spatial space have to be discretized. We therefore introduce discrete points $\{t_i\}_{i=0}^n, t_i = i\Delta_t$ and $\{x_j\}_{j=0}^{m+1}, x_j = x_0 + j\Delta_x, \Delta_x = \frac{x_{\max} - x_{\min}}{m+1} = \frac{x_{m+1} - x_0}{m+1}$. The upper and lower x boundaries $x_{\max} = x_{m+1}, x_{\min} = x_0$ are considered only as hypothetical values. The scheme will *never touch* these hypothetical boundaries.¹⁷⁹ The next step is to deployment of the finite differences over the whole x -grid. We suggest the arrangement shown in the Figure 38. This arrangement ensures that neither $x_{\max} = x_{m+1}$, nor $x_{\min} = x_0$ and values of V at these points are required. The differences are specified such that in the central points $\widehat{\mathcal{L}}V(t, x_j), j = 2, \dots, m-1$ is defined via the central differences, $\widehat{\mathcal{L}}V(t, x_1)$ is specified via the forward difference and finally $\widehat{\mathcal{L}}V(t, x_m)$ via the backward difference. All differences need to produce an error $\varepsilon = O(\Delta_x^2)$ or smaller. This specification will circumvent the unwanted task of specifying the BCs explicitly. We denote δ_x, δ_{xx} the difference operators approximating the first and the second derivative so that

$$\frac{\partial V_j}{\partial x} = \delta_x V_j + O(\Delta_x^2), \quad \frac{\partial^2 V_j}{\partial x^2} = \delta_{xx} V_j + O(\Delta_x^2).$$

Using δ_x, δ_{xx} we define $\widehat{\mathcal{L}}V_j, j = 1, \dots, m$ as

$$\widehat{\mathcal{L}}V_j = \mu(t, x_j) \delta_x V_j + \frac{1}{2} \sigma^2(t, x_j) \delta_{xx} V_j - r(t, x_j) V_j.$$

Now, it is, however, necessary to give a precise specification of $\widehat{\mathcal{L}}V_j$ in accordance with the Figure 38, that is to specify $\widehat{\mathcal{L}}V_1$ in terms of the *forward* differences, $\widehat{\mathcal{L}}V_j, j = 2, \dots, m-1$ in terms of the *central* differences and $\widehat{\mathcal{L}}V_m$ in terms of the *backward* differences. We borrow these finite differences from (Tavella and Randall, 2000, p. 69, Table 3.1).¹⁸⁰ The first and second finite differences are given as follows:

$$\begin{aligned} \frac{\partial V_j}{\partial x} \approx \delta_x V_j &= \begin{cases} \frac{-3V_j + 4V_{j+1} - V_{j+2}}{2\Delta_x} & \text{for forward difference} \\ \frac{V_{j+1} - V_{j-1}}{2\Delta_x} & \text{for central difference} \\ \frac{3V_j - 4V_{j-1} + V_{j-2}}{2\Delta_x} & \text{for backward difference} \end{cases}, \quad (\text{C.1}) \\ \frac{\partial^2 V_j}{\partial x^2} \approx \delta_{xx} V_j &= \begin{cases} \frac{2V_j - 5V_{j+1} + 4V_{j+2} - V_{j+3}}{\Delta_x^2} & \text{for forward difference} \\ \frac{V_{j-1} - 2V_j + V_{j+1}}{\Delta_x^2} & \text{for central difference} \\ \frac{-V_{j-3} + 4V_{j-2} - 5V_{j-1} + 2V_j}{\Delta_x^2} & \text{for backward difference} \end{cases}. \end{aligned}$$

¹⁷⁸sometimes also named 'infinitesimal generator'.

¹⁷⁹this is an example of the contrast with scheme running under Dirichlet BCs. Under Dirichlet BCs the value of the dependent variable V would have to be specified at $x_{\max} = x_{m+1}, x_{\min} = x_0$.

¹⁸⁰there is a *critical* typo in (Tavella and Randall, 2000, p. 69, Table 3.1) on the 5th row (backward derivative $\frac{\partial u}{\partial x}$). $-4u(x + \Delta x, y)$ should be replaced by $-4u(x - \Delta x, y)$ in the formula.

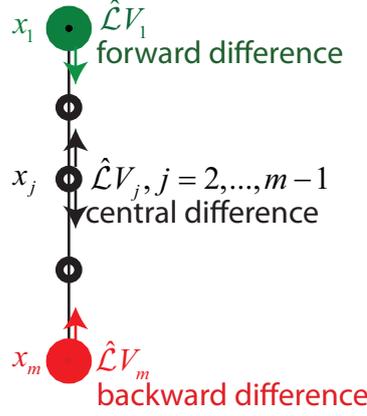


Figure 38: Finite difference scheme in one dimension for any point in time $t_i, i < n$ prior to maturity: The optimal deployment of finite differences on the x -grid.

Using these differences in $\hat{\mathcal{L}}V_1$ we obtain

$$\hat{\mathcal{L}}V_1 = c_1(t)V_1 + u(t)V_2 + uu(t)V_3 + uuu(t)V_4,$$

where

$$\begin{aligned} c_1(t) &= -\frac{3}{2} \frac{\mu(t, x_1)}{\Delta_x} + \frac{\sigma^2(t, x_1)}{\Delta_x^2} - r(t, x_1), & u(t) &= 2 \frac{\mu(t, x_1)}{\Delta_x} - \frac{5}{2} \frac{\sigma^2(t, x_1)}{\Delta_x^2} \\ uu(t) &= -\frac{1}{2} \frac{\mu(t, x_1)}{\Delta_x} + 2 \frac{\sigma^2(t, x_1)}{\Delta_x^2}, & uuu(t) &= -\frac{1}{2} \frac{\sigma^2(t, x_1)}{\Delta_x^2}. \end{aligned}$$

Likewise, for $\hat{\mathcal{L}}V_j, j = 2, \dots, m-1$ we have

$$\hat{\mathcal{L}}V_j = l_j(t)V_{j-1} + c_j(t)V_j + u_j(t)V_{j+1}$$

$$l_j(t) = -\frac{\mu(t, x_j)}{2\Delta_x} + \frac{1}{2} \frac{\sigma^2(t, x_j)}{\Delta_x^2}, \quad c_j(t) = -\frac{\sigma^2(t, x_j)}{\Delta_x^2} - r(t, x_j), \quad u_j(t) = \frac{\mu(t, x_j)}{2\Delta_x} + \frac{1}{2} \frac{\sigma^2(t, x_j)}{\Delta_x^2}.$$

And finally, $\hat{\mathcal{L}}V_m$ reads

$$\hat{\mathcal{L}}V_m = ll(t)V_{m-3} + ll(t)V_{m-2} + l(t)V_{m-1} + c_m(t)V_m$$

$$\begin{aligned} ll(t) &= -\frac{1}{2} \frac{\sigma^2(t, x_m)}{\Delta_x^2}, & ll(t) &= \frac{1}{2} \frac{\mu(t, x_m)}{\Delta_x} + 2 \frac{\sigma^2(t, x_m)}{\Delta_x^2} \\ l(t) &= -2 \frac{\mu(t, x_m)}{\Delta_x} - \frac{5}{2} \frac{\sigma^2(t, x_m)}{\Delta_x^2}, & c_m(t) &= \frac{3}{2} \frac{\mu(t, x_m)}{\Delta_x} + \frac{\sigma^2(t, x_m)}{\Delta_x^2} - r(t, x_m). \end{aligned}$$

The next step is to construct a system of coupled ODEs. Denoting $\mathbf{V}(t) = (V_1(t), \dots, V_m(t))^T$ we can write

$$\hat{\mathcal{L}}\mathbf{V}(t) = \left(\hat{\mathcal{L}}V_1(t), \hat{\mathcal{L}}V_2(t), \dots, \hat{\mathcal{L}}V_m(t) \right)^T = \mathbf{A}(t)\mathbf{V}(t),$$

where

$$\mathbf{A}(t) = \begin{pmatrix} c_1(t) & u(t) & uu(t) & uuu(t) & 0 & \cdots & 0 \\ l_2(t) & c_2(t) & u_2(t) & 0 & 0 & \cdots & 0 \\ 0 & l_3(t) & c_3(t) & u_3(t) & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & l_{m-1}(t) & c_{m-1}(t) & u_{m-1}(t) \\ 0 & \cdots & 0 & ll(t) & ll(t) & l(t) & c_m(t) \end{pmatrix},$$

is a $(m \times m)$ square¹⁸¹ matrix, to capture all the spatial derivatives at a given time t . Now it is possible to use $\widehat{\mathcal{L}}\mathbf{V}(t) = \mathbf{A}(t)\mathbf{V}(t)$ in the space-discretized PDE. We have

$$\begin{aligned} \frac{\partial \mathbf{V}(t)}{\partial t} &= -\widehat{\mathcal{L}}\mathbf{V}(t) - \mathbf{h}(t, \mathbf{x}) + O(\Delta_x^2) \\ &= -\mathbf{A}(t)\mathbf{V}(t) - \mathbf{h}(t, \mathbf{x}) + O(\Delta_x^2). \end{aligned}$$

The only remaining non-discretized derivative is the time derivative $\frac{\partial \mathbf{V}(t)}{\partial t} = \left(\frac{\partial V_1(t)}{\partial t}, \frac{\partial V_2(t)}{\partial t}, \dots, \frac{\partial V_m(t)}{\partial t} \right)^\top$. We apply the Crank-Nicolson¹⁸² and we find

$$\frac{\mathbf{V}(t_{i+1}) - \mathbf{V}(t_i)}{\Delta_t} = -\mathbf{A}(t_i^{i+1})\mathbf{V}(t_i^{i+1}) - \mathbf{h}(t_i^{i+1}, \mathbf{x}) + O(\Delta_x^2) + O(\Delta_t^2),$$

where $t_i^{i+1} = \frac{1}{2}(t_i + t_{i+1})$. After expansion of $\mathbf{A}(t_i^{i+1})\mathbf{V}(t_i^{i+1})$ and rearrangement we have

$$\left(\mathbf{I} - \frac{1}{2}\Delta_t \mathbf{A}(t_i) \right) \mathbf{V}(t_i) = \left(\mathbf{I} + \frac{1}{2}\Delta_t \mathbf{A}(t_{i+1}) \right) \mathbf{V}(t_{i+1}) + \Delta_t \mathbf{h}(t_i^{i+1}, \mathbf{x}) + O(\Delta_x^2) + O(\Delta_t^2),$$

with \mathbf{I} being the identity matrix of the same dimensions like \mathbf{A} .

C.2 Two-dimensional PDEs

The one-dimensional scheme in the Appendix C.1 was quite straightforward and also easy to be implemented. The applications described in this text are, however, typically dealing with two-dimensional PDEs. Frequently, one factor is the risk-free rate $x_1 = r$ and the second factor is the default intensity $x_2 = \gamma$, or any transformation of it. We show a scheme analogous to the one-dimensional case, which is, however, *significantly* more difficult for the implementation because there is an extra mixed-derivative term which

¹⁸¹notice that the matrix \mathbf{A} is *not* tri-diagonal, while it would be tri-diagonal if we used Dirichlet BCs. In turn its inversion is slightly more demanding but there are three points to consider. (i) the matrix \mathbf{A} is often time-homogeneous (or time invariant) hence the inversion takes place only once. (ii) as the boundary conditions are exact, compared to Dirichlet, where the BCs are usually only approximative, less points on the grid are needed. This speeds the up the inversion. (iii) the scheme operates with error $O(\Delta_x^2)$ and this again allows to achieve good results with only a few spatial points on the grid.

¹⁸²this corresponds to setting $\theta = 1/2$ in a classial θ -scheme.

makes the scheme much more complex. Assume a boundary value problem of the kind

$$\begin{aligned} \frac{\partial V}{\partial t} + \mu_1(t, x) \frac{\partial V}{\partial x_1} + \mu_2(t, x) \frac{\partial V}{\partial x_2} + \frac{1}{2} \sigma_1^2(t, x) \frac{\partial^2 V}{\partial x_1^2} + \frac{1}{2} \sigma_2^2(t, x) \frac{\partial^2 V}{\partial x_2^2} \\ + \sigma_1(t, x) \sigma_2(t, x) \rho(t, x) \frac{\partial^2 V}{\partial x_1 \partial x_2} + h(t, x) - r(t, x) V = 0, \end{aligned} \quad (\text{C.2})$$

with the terminal condition $V(t, x) = g(x)$. We introduce the following infinitesimal operators

$$\begin{aligned} \mathcal{L}_i &= \mu_i(t, x) \frac{\partial}{\partial x_i} + \frac{1}{2} \sigma_i^2(t, x) \frac{\partial^2}{\partial x_i^2} - \frac{1}{2} r(t, x), \text{ for } i = 1, 2, \\ \mathcal{L}_{1,2} &= \sigma_1(t, x) \sigma_2(t, x) \rho(t, x) \frac{\partial^2}{\partial x_1 \partial x_2}, \end{aligned}$$

such that the PDE (C.2) in terms of $\mathcal{L}_1, \mathcal{L}_2$ and $\mathcal{L}_{1,2}$ is compactly

$$\frac{\partial V}{\partial t} + \mathcal{L}_1 V + \mathcal{L}_2 V + \mathcal{L}_{1,2} V + h(t, x) = 0. \quad (\text{C.3})$$

The partials in $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_{1,2}$ now have to be approximated by finite differences (resulting into differences-based operators $\hat{\mathcal{L}}_1, \hat{\mathcal{L}}_2, \hat{\mathcal{L}}_{1,2}$). In order to maintain the second-order error across the whole grid, we suggest the arrangement of the partials as shown in the Figure 39. The first and the second differences (forward, central and backward) approximating $\frac{\partial V}{\partial x_i}$

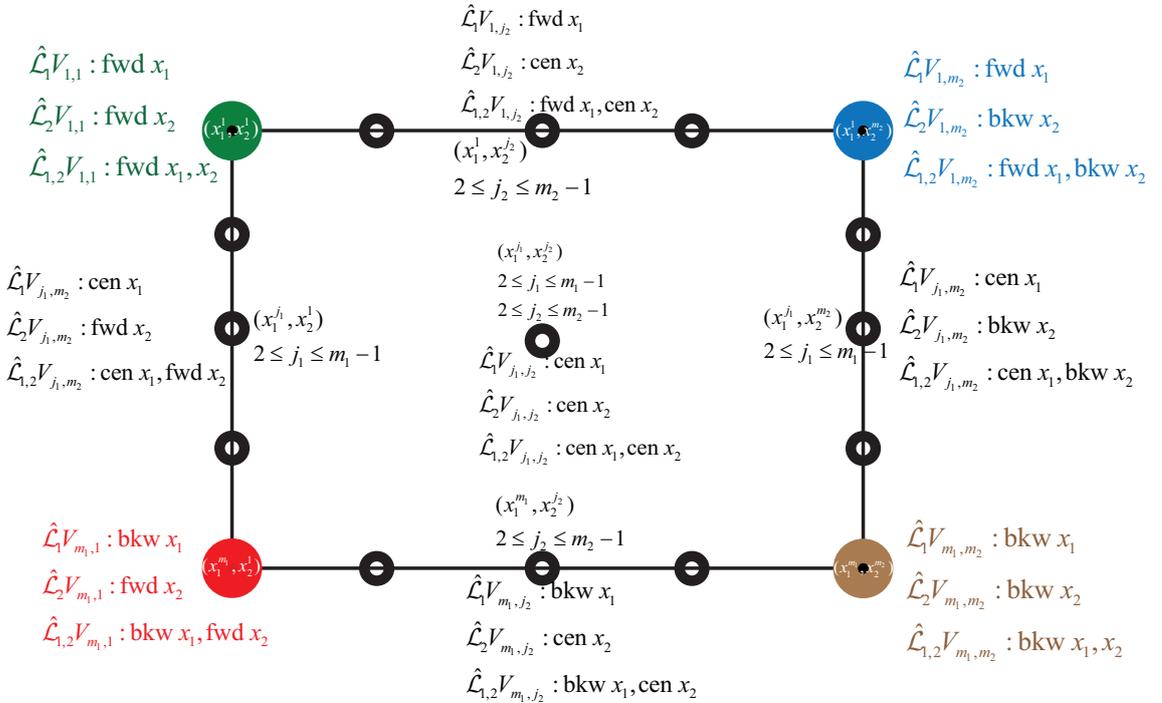


Figure 39: The optimal deployment of differences across a two-dimensional grid.

and $\frac{\partial^2 V}{\partial x_i^2}$ with second-order errors can be taken from the Appendix C.1. A more challenging

task is to approximate the *mixed* partial $\frac{\partial^2 V}{\partial x_1 \partial x_2}$ with the second-order precision in all the points in the Figure 39. In order to derive these finite differences, we follow (Zikanov, 2010, Section 4), who simply suggests to take consecutive differences of a given error order in x and y . To illustrate the method of taking consecutive differences, let us work with some general function f (instead of V) and spatial variables x, y instead of x_1 and x_2 , respectively. Assume that we want to find the finite differences approximations to $\frac{\partial^2 f}{\partial x \partial y}$ such that these differences are forward in the x -direction and backward in the y -direction and the total spatial error is $\varepsilon = O((\Delta x)^2) + O((\Delta y)^2)$.¹⁸³ We first take the first forward difference in x of the order $O((\Delta x)^2)$ and so based on (C.1) we have

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) = \frac{-3 \frac{\partial f(x,y)}{\partial y} + 4 \frac{\partial f(x+\Delta x,y)}{\partial y} - \frac{\partial f(x+2\Delta x,y)}{\partial y}}{2\Delta x} + O((\Delta x)^2). \quad (\text{C.4})$$

Now it is necessary to approximate each of the y -partials with the first backward differences, maintaining $O((\Delta y)^2)$. We have (again using (C.1), yet for the variable y)

$$\begin{aligned} \frac{\partial f(x,y)}{\partial y} &= \frac{3f(x,y) - 4f(x,y - \Delta y) + f(x,y - 2\Delta y)}{2\Delta x} + O((\Delta y)^2) \\ \frac{\partial f(x + \Delta x,y)}{\partial y} &= \frac{3f(x + \Delta x,y) - 4f(x + \Delta x,y - \Delta y) + f(x + \Delta x,y - 2\Delta y)}{2\Delta x} + O((\Delta y)^2) \\ \frac{\partial f(x + 2\Delta x,y)}{\partial y} &= \frac{3f(x + 2\Delta x,y) - 4f(x + 2\Delta x,y - \Delta y) + f(x + 2\Delta x,y - 2\Delta y)}{2\Delta x} + O((\Delta y)^2). \end{aligned}$$

Substitution of these terms back into (C.4) finally yields

$$\begin{aligned} \frac{\partial^2 f}{\partial x \partial y} &= \frac{1}{4\Delta x \Delta y} (-9f(x,y) + 12f(x,y - \Delta y) - 3f(x,y - 2\Delta y)) \\ &+ \frac{1}{4\Delta x \Delta y} (12f(x + \Delta x,y) - 16f(x + \Delta x,y - \Delta y) + 4f(x + \Delta x,y - 2\Delta y)) \\ &+ \frac{1}{4\Delta x \Delta y} (-3f(x + 2\Delta x,y) + 4f(x + 2\Delta x,y - \Delta y) - f(x + 2\Delta x,y - 2\Delta y)) \\ &+ O((\Delta x)^2) + O((\Delta y)^2). \end{aligned}$$

We compactly write the outcome as a triplet

$\{(-9, 12, -3, 12, -16, 4, -3, 4, -1), (0, 0, 0, 1, 1, 1, 2, 2, 2), (0, -1, -2, 0, -1, -2, 0, -1, -2)\}$.

The complete list of the finite differences needed to approximate $\frac{\partial^2 f(x,y)}{\partial x \partial y}$ (or $\frac{\partial^2 V}{\partial x_1 \partial x_2}$ in our pricing problems) in the scheme presented in the Figure 39 can be found in the Table 21.

Assume that the form of the finite differences at all points of the scheme has been determined in accordance with the Figure 39 so that $\widehat{\mathcal{L}}_1 V_{j_1, j_2}, \widehat{\mathcal{L}}_2 V_{j_1, j_2}, \widehat{\mathcal{L}}_{1,2} V_{j_1, j_2}$ are known at *all* points $j_1 = 1, \dots, m_1$ and $j_2 = 1, \dots, m_2$. Using Crank-Nicolson, we set up a discrete-form of (C.3) which is

$$\begin{aligned} \frac{V_{j_1, j_2}(t_{i+1}) - V_{j_1, j_2}(t_i)}{\Delta t} &= - \left(\frac{1}{2} \widehat{\mathcal{L}}_1 V_{j_1, j_2}(t_{i+1}) + \frac{1}{2} \widehat{\mathcal{L}}_1 V_{j_1, j_2}(t_i) + \frac{1}{2} \widehat{\mathcal{L}}_2 V_{j_1, j_2}(t_{i+1}) \right. \\ &\quad \left. + \frac{1}{2} \widehat{\mathcal{L}}_2 V_{j_1, j_2}(t_i) + \frac{1}{2} \widehat{\mathcal{L}}_{1,2} V_{j_1, j_2}(t_{i+1}) + \frac{1}{2} \widehat{\mathcal{L}}_{1,2} V_{j_1, j_2}(t_i) \right) - h(t_i^{i+1}, x) + \varepsilon, \end{aligned}$$

¹⁸³this corresponds to the mixed derivative $\frac{\partial^2 V}{\partial x_1 \partial x_2}$ at the point $(x_1^1, x_2^{m_2})$ in the Figure 39.

x dir	y dir	f coefficients	Δx coefficients	Δy coefficients
fwd	fwd	9,-12,3,-12,16,-4,3,-4,1	0,0,0,1,1,1,2,2,2	0,1,2,0,1,2,0,1,2
bkw	bkw	1,-4,3,-4,16,-12,3,-12,9	-2,-2,-2,-1,-1,-1,0,0,0	-2,-1,0,-2,-1,0,-2,-1,0
bkw	fwd	-3,4,-1,12,-16,4,-9,12,-3	-2,-2,-2,-1,-1,-1,0,0,0	0,1,2,0,1,2,0,1,2
cen	fwd	-3,4,-1,3,-4,1	1,1,1,-1,-1,-1	0,1,2,0,1,2
cen	bkw	1,-4,3,-1,4,-3	1,1,1,-1,-1,-1	-2,-1,0,-2,-1,0
cen	cen	1,-1,-1,1	1,1,-1,-1	1,-1,1,-1

Table 21: Coefficients in the finite differences accurate to $O((\Delta x)^2) + O((\Delta y)^2)$ approximating the mixed derivative $\frac{\partial^2 f(x,y)}{\partial x \partial y}$. The denominator in all cases is $4\Delta x \Delta y$. To find other derivatives, only switch the columns 'Δx coefficients' and 'Δy coefficients' of a mirror derivative. For example (fwd x , bkw y) can be obtained from (bkw x , fwd y).

where $x = (x_1^{j_1}, x_2^{j_2})$ and $\varepsilon = O(\Delta_1^2) + O(\Delta_2^2) + O(\Delta_t^2)$. After grouping the variables according to the time variable, we have

$$V_{j_1, j_2}(t_i) - \Delta_t \frac{1}{2} (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2 + \widehat{\mathcal{L}}_{1,2}) V_{j_1, j_2}(t_i) = V_{j_1, j_2}(t_{i+1}) + \Delta_t \frac{1}{2} (\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2 + \widehat{\mathcal{L}}_{1,2}) V_{j_1, j_2}(t_{i+1}) + \Delta_t h(t_i^{i+1}, \mathbf{x}) + \varepsilon.$$

This can be rearranged into a more usable matrix form

$$\left(\mathbf{I} - \Delta_t \frac{1}{2} \mathbf{A}(t_i) \right) \mathbf{V}(t_i) = \left(\mathbf{I} + \Delta_t \frac{1}{2} \mathbf{A}(t_{i+1}) \right) \mathbf{V}(t_{i+1}) + \Delta_t \mathbf{h}(t_i^{i+1}, \mathbf{x}) + \varepsilon, \quad (\text{C.5})$$

where \mathbf{I} is $(m_1 m_2 \times m_1 m_2)$ identity matrix, $\mathbf{A}(t)$ is $(m_1 m_2 \times m_1 m_2)$ matrix of coefficients from $\widehat{\mathcal{L}}_1 + \widehat{\mathcal{L}}_2 + \widehat{\mathcal{L}}_{1,2}$ that apply to a $m_1 m_2$ -dimensional (column) vector $\mathbf{V}(t)$. The biggest technical challenge is to organize the terms in all $\mathbf{V}, \mathbf{A}, \mathbf{h}$ with \mathbf{A} being the most complicated part.

Although there are many ways of the arrangement of \mathbf{V} , we suggest

$$\mathbf{V}(t) = (V_{1,1}(t), V_{2,1}(t), \dots, V_{m_1,1}(t), V_{1,2}(t), \dots, V_{m_1,2}(t), \dots, V_{1,m_1}(t), \dots, V_{m_1,m_2}(t))^T.$$

This implies that $V_{j_1, j_2}(t)$ is $(j_2 - 1)m_1 + j_1$ th element of $\mathbf{V}(t)$. To best understand \mathbf{A} in the context of the PDE (C.3) it suffices to recall that at a given time-point the spatial derivatives captured by $\mathcal{L}_1 V, \mathcal{L}_2 V$, and $\mathcal{L}_{1,2} V$ need to be evaluated at every point in the space (grid). In turn, a row $(j_2 - 1)m_1 + j_1$ of \mathbf{A} holds the difference (differential) coefficients from $\widehat{\mathcal{L}}_1 V_{j_1, j_2}, \widehat{\mathcal{L}}_2 V_{j_1, j_2}, \widehat{\mathcal{L}}_{1,2} V_{j_1, j_2}$, when we evaluate the discretized PDE at the point $(x_1^{j_1}, x_2^{j_2})$ in space. The position (element) of a coefficient a fixed row of \mathbf{A} then indicates which V from \mathbf{V} the coefficient multiplies according to the specification of the finite difference. The vector $\mathbf{h}(t_i^{i+1}, \mathbf{x})$ is simply

$$\mathbf{h}(t_i^{i+1}, \mathbf{x}) = (h(t_i^{i+1}, x_1^1, x_2^1), \dots, h(t_i^{i+1}, x_1^{m_1}, x_2^1), h(t_i^{i+1}, x_1^1, x_2^2), \dots)^T,$$

that is, the ordering of the indices j_1, j_2 of $(x_1^{j_1}, x_2^{j_2})$ must be same as in $\mathbf{V}(t)$. The filling procedure of \mathbf{A} is best described algorithm-wise. We show this in the Box C.2.

Box C.1: Pseudocode: filling of \mathbf{A}

```

% create (m1*m2 X m1*m2) matrix A of zeros
A = zeros(m1*m2);
% iterate over every point in the space (grid)
for j1 = 1:m1
    for j2 = 1:m2
        % for (j1, j2) get the row index of the A matrix
        % this row will be then filled by the coefficients
        A_row_idx = (j2-1)*m1+j1;

        % compute vectors of coeffs resulting from L1, L2, L12 operators
        % also get vectors of positions (columns of A) of these coeffs
        [coeffs1, coeffs_col1] = lcu1(j1, j2, x1_grid, x2_grid, dx1, ..);
        [coeffs2, coeffs_col2] = lcu2(j1, j2, x1_grid, x2_grid, dx2, ..);
        [coeffs12, coeffs_col12] = lcu12(j1, j2, x1_grid, x2_grid, dx1*dx2,..);

        % save the coefficients to row A_row_idx of A
        % notice the coefficients may add up (A = A + coeffs)
        % leave zeros everywhere else on the row A_row_idx of A
        A(A_row_idx, coeffs_col1) = A(A_row_idx, coeffs_col1) + coeffs1;
        A(A_row_idx, coeffs_col2) = A(A_row_idx, coeffs_col2) + coeffs2;
        A(A_row_idx, coeffs_col12) = A(A_row_idx, coeffs_col12) + coeffs12;
    end
end
end

```

The sparsity pattern of \mathbf{A} is visualized in the Figure 40 and it is easy to see that the matrix is not a (tri- or penta-)diagonal matrix. Notice in the Box C.2 the critical role of the functions `lcu1`, `lcu2` and `lcu12`. Each of these functions delivers a set of coefficients (and their positions in a row of \mathbf{A}) in $\widehat{\mathcal{L}}_1 V_{j_1, j_2}$, $\widehat{\mathcal{L}}_2 V_{j_1, j_2}$, and $\widehat{\mathcal{L}}_{1,2} V_{j_1, j_2}$, respectively. In the body of `lcu1`, `lcu2` and `lcu12` it must be checked *where on the grid* the derivatives (differences) in $\widehat{\mathcal{L}}$ are actually computed according to the values (j_1, j_2) and an appropriate set of coefficients and their positions in the row of \mathbf{A} returned. This choice of derivatives according to the position (j_1, j_2) is dictated by the Figure 39. In fact the choice of derivatives in `lcu1` ($\widehat{\mathcal{L}}_1$) is only determined by j_1 , in `lcu2` ($\widehat{\mathcal{L}}_2$) the choice of derivatives is only determined by j_2 but in `lcu12` ($\widehat{\mathcal{L}}_{1,2}$) the choice of derivatives is determined by *both* j_1 and j_2 . `lcu12` is thus a much more complicated function than the other two, as many variants of derivatives have to be carried out inside. This implementation through the 'separation' of the differential operators, however, is very straightforward to follow.

In the most situations, when we iterate backwards in time, the system (C.5) is *not* solved through the matrix inversion but rather via iterative solvers of systems of linear equations. This is because the matrix \mathbf{A} for 2D (and other multi-dimensional) PDEs is often very large and non (tri- or penta-)diagonal which makes it very difficult to invert. Additionally, if

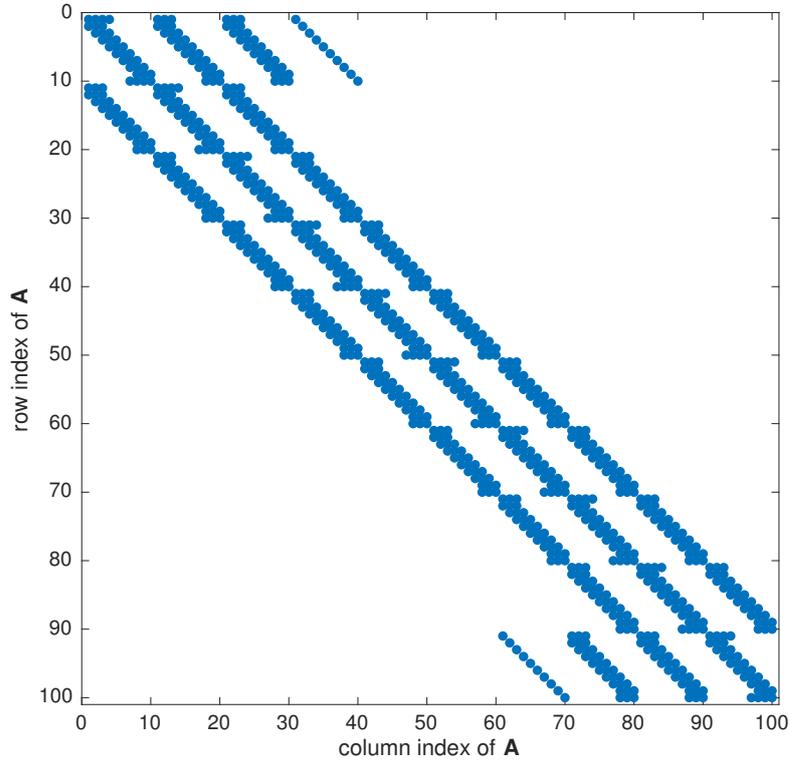


Figure 40: The sparsity pattern of \mathbf{A} with $m_1 = m_2 = 10$ (hence a 100×100 matrix).

furthermore the matrix depends on time¹⁸⁴ ($\mathbf{A}(t) \neq \mathbf{A}$), then it would have to be inverted at every time step. Since \mathbf{A} has a high degree of sparsity (see the Figure 40) the solution to the system of linear equation can be efficiently solved by various iterative solvers. We conducted a particularly intensive numerical test of various solvers and the outcomes are shown in the Table 22. When we dismiss all the solvers (`minres`, `pcg` and `symmlq`) unable to at least approximately match the true price of the sample instrument, then there are still many algorithms which are substantially faster (yet accurate) than the standard matrix inversion. We found the *Biconjugate gradients stabilized* method (`bicgstab`) to be the best available method overall. The total win of the linear system solvers over the matrix inversion can easily be justified by the fact that many of these methods largely benefit from the matrix sparsity while the matrix inversion takes no advantage of the matrix sparsity. Because $\mathbf{A}(t)$ (or \mathbf{A}) is a highly sparse matrix (see again the Figure 40) the linear system solvers are in turn very efficient.

¹⁸⁴as a practical example assume a pricing problem of a defaultable bond (subject to some default intensity $x_2 = \gamma$) and a risk-free rate $x_1 = r$ following the *Hull-White's* dynamics. Then the PDE for the bond price has time-dependent coefficient matrix, that is $\mathbf{A}(t) \neq \mathbf{A}$. This is necessarily given by the time dependency of the drift term in the Hull-White model and possibly (but less commonly) also by some time dependency of the intensity process.

method	MATLAB function	time (s)	price
<i>plain matrix inversion</i>	<code>inv</code>	48.0215	0.8707
Biconjugate gradients	<code>bicg</code>	0.1868	0.8707
Biconjugate gradients stabilized	<code>bicgstab</code>	0.1460	0.8707
Biconjugate gradients stabilized (I)	<code>bicgstabl</code>	0.2640	0.8707
Conjugate gradients squared	<code>cgs</code>	0.1680	0.8707
Generalized minimum residual	<code>gmres</code>	0.2429	0.8707
Minimum residual	<code>minres</code>	0.2505	0.0029
Preconditioned conjugate gradients	<code>pcg</code>	0.2102	0.0548
Quasi-minimal residual	<code>qmr</code>	0.2348	0.8706
Sparse least-squares	<code>lsqr</code>	0.3721	0.8722
Symmetric LQ	<code>symmlq</code>	0.2457	0.0650
Transpose-free quasi-minimum residual	<code>tfqmr</code>	0.2175	0.8707

Table 22: Comparison of tested linear system solvers on a 2D PDE scheme with $m_1 = m_2 = 100$ ($\mathbf{A}(t) = \mathbf{A}$ of 100 000 000 elements). The first row (*plain matrix inversion*) serves as a benchmark. The priced claim was a defaultable bond with a correct price 0.8707.

D Estimation of SDE/JDE parameters using PDE/PIDE

In many problems, we are dealing with SDEs or JDEs which describe the dynamics of processes. These equations are subject to parameters that can either be obtained by fitting to quoted prices or possibly estimated from time series. Which method of the calibration is selected depends more on the later use rather than on our choice. Here, we describe a method for the estimation of parameters in SDE/JDE from observed time series. Using the standard methods such as MLE this is often a delicate task as a complicated SDE/JDE might not have a closed-form objective (log)likelihood function. Here, we introduce a novel, general, PDE/PIDE-based estimation method allowing to extract parameters for a (correlated) system of SDEs/JDEs (also with possible simultaneous jumps) which could be in practice very difficult. The method has the following advantages:

- it is based on a natural, differential form of SDE/JDE (that is just $dx(t)$) and does *not* need a formula for its solution ($x(t)$). The majority of equations of dynamics are formulated using SDEs/JDEs,
- the observations can be freely dispersed in time. One such example are illiquid instruments whose price information arrive irregularly in time,
- conditional on equal grid and time spacing it is easy to compare fit of various dynamics to the time series. One can thus easily conclude which specification of JDE/SDE fits the data series the best.

Let us assume that the dynamics of the process x is subject to the SDE

$$dx(t) = \mu(t, x(t))dt + \sigma(t, x(t))dW(t),$$

where all the components are scalar and the parameter set in the differential $dx(t)$ is compactly denoted Θ . We also assume that we observe a realized time series $x_0^O, x_1^O, \dots, x_n^O$ of the process x at times t_0, t_1, \dots, t_n . The time spacing between the observations is denoted $\Delta t_i = t_i - t_{i-1}, i = 1, \dots, n$ and is not necessarily equidistant. Let $\varepsilon > 0$ be a 'small', strictly positive constant and define

$$g_i(x) = \mathbf{1}_{\{x \in [x_i^O - \varepsilon, x_i^O + \varepsilon]\}}.$$

If we now form the conditional expectation

$$\mathbb{E} [g_i(x(t_i)) | x(t_{i-1}) = x_{i-1}^O], \quad (\text{D.1})$$

we find that this is in fact equal to the (conditional) transition probability

$$P_i = \mathbb{P} [x(t_i) \in [x_i^O - \varepsilon, x_i^O + \varepsilon] | x(t_{i-1}) = x_{i-1}^O] = \mathbb{E} [g_i(x(t_i)) | x(t_{i-1}) = x_{i-1}^O].$$

Assume further a time frame $[t_{i-1}, t_i]$ and restrict the time variable t to be in this time frame, so that $t \in [t_{i-1}, t_i]$ and further assume $P_i = P_i(t, x)$. Since the dynamics of x is that of Itô, the expectation (D.1) satisfies Feynman-Kac and so necessarily P_i solves

$$\frac{\partial P_i}{\partial t} + \mu(t, x) \frac{\partial P_i}{\partial x} + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 P_i}{\partial x^2} = 0,$$

with the terminal condition at $t = t_i$ defined as

$$P_i(t_i, x) = g_i(x) = \mathbf{1}_{\{x \in [x_i^O - \varepsilon, x_i^O + \varepsilon]\}}.$$

Because the transition probability P_i is also determined by the parameter set Θ , we write it as $P_i(\Theta) = P_i(t, x | \Theta)$. As the primary objective is to find the set of parameters $\hat{\Theta}$ that best describe the trajectory in terms of the probability, the remainder of the approach is an obvious MLE analogy and the estimator is equal to

$$\hat{\Theta} = \arg \min_{\Theta} \left[- \sum_{i=1}^n \ln P_i(\Theta) \right].$$

Let us allow a few comments to this method. Interestingly, the approach perceives ' $dt = \Delta t_i$ ' as not an infinitesimally short horizon but a rather some meaningful positive horizon that is zoomed in and can further be discretized (in the PDE scheme). That is $x(t), t \in [t_{i-1}, t_i]$ is free to evolve as if the horizon were not short. Secondly, we might stumble upon the choice of ε which in fact defines what is the tolerance of the modeled variable $x(t_i)$ to be 'sufficiently close' to the observation x_i^O so that it is regarded as a success. Without any grid refinements ε set too large could mean that if x arrives at several states at $t = t_i$ all these arrivals count as success (= 1). However, this is probabilistically completely incorrect as arrival to only one, unique, state can occur. Setting ε too low might then on the other hand result into the pathological situation that x would never hit x_i^O which is also incorrect. To circumvent these issues with the choice of ε we suggest to arrange the grid so that one of the points on the grid lies exactly on x_i^O . Then, with the abuse of the above stated assumption about ε being strictly positive, we conclude that we allow x to arrive only into finitely many 'terminal' states, one of which

is exactly x_i^O and so $\varepsilon = 0$. Finally, we should comment the boundary conditions. These are trivial and without loss of generality, they can be set to $P_i(t, x_{\min}) = P_i(t, x_{\max}) = 0$. This is because the transition density goes to 0 if we consider extreme movements in a short period Δt_i .

These ideas, however, can be pushed further and the general result will still hold. In particular assuming x is a d -dimensional stochastic process (possibly with pure jumps) with dynamics defined by a stochastic differential

$$dx(t) = \mu(t, x(t))dt + \sigma(t, x(t))dW(t) + dJ(t, x(t)),$$

this means the terminal 'success' function is now

$$g_i(x) = \mathbf{1}_{\{x_1 \in [x_{1,i}^O - \varepsilon_1, x_{1,i}^O + \varepsilon_1], x_2 \in [x_{2,i}^O - \varepsilon_2, x_{2,i}^O + \varepsilon_2], \dots\}},$$

and P_i solves PDE (or PIDE if jumps are involved)

$$\frac{\partial P_i}{\partial t} + \sum_{j=1}^d \mathcal{A}_{c,j} P_i + \mathcal{A}_c P_i \text{ of mixed terms} + \sum_{j=1}^d \mathcal{A}_{J,j} P_i + \mathcal{A}_J P_i \text{ of simultaneous jumps} = 0,$$

where \mathcal{A}_c are compactly infinitesimal generators of the diffusion and \mathcal{A}_J are infinitesimal generators of the jump components. The terminal condition obviously remains the same and it reads

$$P_i(t_i, x) = g_i(x).$$

E Proof of the Black-Scholes model with Hull-White's r

Proof. For the proof of the result (II.1.15) we will borrow the general option formula from (Geman et al., 1995). For simplicity, we assume away the dividends q .

To start, we take as given that the Hull-White model in the HJM framework (see the Appendix A.2) is uniquely defined via the forward-rate \mathbb{Q} -dynamics

$$df(t, T) = (\dots) dt + \sigma_r e^{-\kappa(T-t)} dW_1^{\mathbb{Q}}(t).$$

The Proposition A.1 immediately gives the \mathbb{Q} -measure bond price dynamics from df and so

$$dB(t, T) = r(t)B(t, T)dt + B(t, T)\sigma_r(t, T)dW_1^{\mathbb{Q}}(t), \text{ where } \sigma_r(t, T) = \frac{\sigma_r}{\kappa} \left(e^{-\kappa(T-t)} - 1 \right).$$

The dynamics of both X and $B(\cdot, T)$ under \mathbb{Q} is thus completely known.

Considering a call option C with a strike K on an asset X , (Geman et al., 1995) argues that the general formula for $C(t, T)$ is equal to

$$C(t, T) = X(t)\mathbb{S}[X(T) \geq K] - KB(t, T)\mathbb{Q}_T[X(T) \geq K], \quad (\text{E.1})$$

where \mathbb{S} is the stock measure¹⁸⁵ and \mathbb{Q}_T is the T -forward measure. We will firstly solve the second probability term $\mathbb{Q}_T [X(T) \geq K]$. Let us set $Z(t) = \frac{X(t)}{B(t,T)}$ and use it in the probability term $\mathbb{Q}_T [\cdot]$ so that

$$\mathbb{Q}_T [X(T) \geq K] = \mathbb{Q}_T \left[\frac{X(T)}{B(T,T)} \geq K \right] = \mathbb{Q}_T [Z(T) \geq K].$$

In turn, the dynamics of Z under \mathbb{Q}_T is needed. The idea is to find Z under \mathbb{Q} and change the measure to \mathbb{Q}_T . The dynamics of Z under \mathbb{Q} can be obtained by several applications of Itô's lemma. This involves first computing

$$d \left(\frac{1}{B(t,T)} \right) = -\frac{1}{B(t,T)} (r(t)dt + \sigma_r(t,T)dW_1(t) - \sigma_r^2(t,T)dt).$$

Now the required differential of the bond-deflated stock process Z is

$$dZ(t) = d \left(\frac{X(t)}{B(t,T)} \right) = d \left(X(t) \frac{1}{B(t,T)} \right),$$

and can be computed using the Itô's product rule (see Corollary I.1.2). We find

$$\begin{aligned} dZ(t) &= \frac{1}{B(t,T)} dX(t) + X(t) \left(-\frac{1}{B(t,T)} (r(t)dt + \sigma_r(t,T)dW_1^\mathbb{Q}(t) - \sigma_r^2(t,T)dt) \right) \\ &+ dX(t) \left(-\frac{1}{B(t,T)} (r(t)dt + \sigma_r(t,T)dW_1^\mathbb{Q}(t) - \sigma_r^2(t,T)dt) \right) \\ &\vdots \\ &= Z(t) \left(\sigma_x(\rho dW_1^\mathbb{Q}(t) + \sqrt{1-\rho^2}dW_2^\mathbb{Q}(t)) - \sigma_r(t,T)dW_1^\mathbb{Q}(t) + \sigma_r^2(t,T)dt \right) \\ &- Z(t) \left(\sigma_r(t,T)dW_1^\mathbb{Q}(t) \right) \left(\sigma_x(\rho dW_1^\mathbb{Q}(t) + \sqrt{1-\rho^2}dW_2^\mathbb{Q}(t)) \right). \end{aligned}$$

At this point, we can expand the terms such the differential is significantly reduced by virtue of the quadratic variation property (here in the form of the multiplication of the differentials; see also the Section I.1.2). We have

$$dZ(t) = Z(t) \left(\sigma_x(\rho dW_1^\mathbb{Q}(t) + \sqrt{1-\rho^2}dW_2^\mathbb{Q}(t)) - \sigma_r(t,T)dW_1^\mathbb{Q}(t) + \sigma_r^2(t,T)dt - \sigma_r(t,T)\sigma_x\rho dt \right).$$

Regrouping and collecting the $dt, dW_1^\mathbb{Q}(t), dW_2^\mathbb{Q}(t)$ -terms gives

$$dZ(t) = Z(t) (\dots) dt + Z(t) \left((\sigma_x\rho - \sigma_r(t,T)) dW_1^\mathbb{Q}(t) + \sigma_x\sqrt{1-\rho^2}dW_2^\mathbb{Q}(t) \right).$$

Since Z is actually X deflated by the T -bond $B(\cdot, T)$, it must be a martingale under the T -forward measure \mathbb{Q}_T and so in view of the Martingale Representation Theorem I.1.1, Z follows

$$dZ(t) = Z(t) \left((\sigma_x\rho - \sigma_r(t,T)) dW_1^{\mathbb{Q}_T}(t) + \sigma_x\sqrt{1-\rho^2}dW_2^{\mathbb{Q}_T}(t) \right),$$

under \mathbb{Q}_T . This can be more compactly written in terms of vectors as

$$\begin{aligned} dZ(t) &= Z(t)\sigma_Z(t)dW^{\mathbb{Q}_T}(t), \text{ where} \\ \sigma_Z(t) &= \left(\sigma_x\rho - \sigma_r(t,T), \sigma_x\sqrt{1-\rho^2} \right), \text{ and } dW^{\mathbb{Q}_T}(t) = \begin{pmatrix} dW_1^{\mathbb{Q}_T}(t) \\ dW_2^{\mathbb{Q}_T}(t) \end{pmatrix}. \end{aligned}$$

¹⁸⁵here, in the sense of the 'asset measure'.

Because Z is an exponential martingale, its solution is

$$Z(T) = \frac{X(t)}{B(t, T)} \exp \left(-\frac{1}{2} \int_t^T |\sigma_Z(u)|^2 du + \int_t^T \sigma_Z(u) dW^{\mathbb{Q}_T}(u) \right).$$

As σ_Z is a deterministic function of time, the stochastic integral is actually a normally distributed random variable with zero mean and variance

$$v_{\text{HW}}(t, T) = \int_t^T |\sigma_Z(u)|^2 du, \quad (\text{E.2})$$

so in the exponent of $Z(T)$ we have a random variable

$$\omega \sim N \left(-\frac{1}{2} v_{\text{HW}}(t, T), v_{\text{HW}}(t, T) \right).$$

A standard machinery then shows that

$$\mathbb{Q}_T [X(T) \geq K] = N(d_2), \quad \text{where } d_2 = \frac{\ln \left(\frac{X(t)}{KB(t, T)} \right) - \frac{1}{2} v_{\text{HW}}(t, T)}{\sqrt{v_{\text{HW}}(t, T)}}. \quad (\text{E.3})$$

Because the 'overall instantaneous volatility' σ_Z is given through the instantaneous variance

$$|\sigma_Z(u)|^2 = (\sigma_x \rho - \sigma_r(u, T))^2 + \left(\sigma_x \sqrt{1 - \rho^2} \right)^2 = \sigma_x^2 - 2\sigma_x \rho \sigma_r(u, T) + \sigma_r^2(u, T),$$

the 'total variance' $v_{\text{HW}}(t, T)$ accumulated between t and T can be computed using the simple integration (E.2) and it can be easily checked that it corresponds to the form (II.1.15).

The solution to the probability term $\mathbb{S}[X(T) \geq K]$ can be processed nearly analogously. It is, however, useful to rewrite it as

$$\mathbb{S}[X(T) \geq K] = \mathbb{S} \left[\frac{1}{X(T)} \leq \frac{1}{K} \right] = \mathbb{S} \left[\frac{B(T, T)}{X(T)} \leq \frac{1}{K} \right] = \mathbb{S} \left[\frac{1}{Z(T)} \leq \frac{1}{K} \right] = \mathbb{S} \left[Y(T) \leq \frac{1}{K} \right],$$

where a new process $Y(t) = \frac{1}{Z(t)} = \frac{B(t, T)}{X(t)}$ was defined. Similar steps as above can be taken to find the differential $dY(t)$ under the stock measure \mathbb{S} . This again gives the exponential martingale

$$dY(t) = Y(t) \left((\sigma_r(t, T) - \sigma_x \rho) dW_1^{\mathbb{S}}(t) - \sigma_x \sqrt{1 - \rho^2} dW_2^{\mathbb{S}}(t) \right),$$

with the solution

$$Y(T) = \frac{B(t, T)}{X(t)} \exp \left(-\frac{1}{2} \int_t^T |\sigma_Y(u)|^2 du + \int_t^T \sigma_Y(u) dW^{\mathbb{S}}(u) \right),$$

where $W^{\mathbb{S}}$ is a column Wiener process under \mathbb{S} and

$$\sigma_Y(t) = \left(\sigma_r(t, T) - \sigma_x \rho, -\sigma_x \sqrt{1 - \rho^2} \right).$$

Similarly to the Z -case, in the exponent of Y there is a normally distributed random variable with mean $-\frac{1}{2}v_{\text{HW}}(t, T)$ and variance $v_{\text{HW}}(t, T)$. Using these properties in $\mathbb{S}\left[Y(T) \leq \frac{1}{K}\right]$ yields

$$\mathbb{S}[X(T) \geq K] = \mathbb{S}\left[Y(T) \leq \frac{1}{K}\right] = N(d_1), \text{ where } d_1 = \frac{\ln\left(\frac{X(t)}{KB(t, T)}\right) + \frac{1}{2}v_{\text{HW}}(t, T)}{\sqrt{v_{\text{HW}}(t, T)}}, \quad (\text{E.4})$$

and so we find $d_2 = d_1 - \sqrt{v_{\text{HW}}(t, T)}$. Using (E.3) and (E.4) in (E.1) we have proved the result (II.1.15). \square

References

- Albrecher, H., Mayer, P., and Schoutens, W. T. 2007. The little Heston trap. *Wilmott Magazine, January issue*, pages 83–92.
- Andersen, L. and Piterbarg, V. V. 2010. Interest rate modeling, volume I: Foundations and vanilla models. *Atlantic Financial Press*, 16:33–43.
- Baran, J. 2016. Post-crisis valuation of derivatives. *Ph.D. thesis, University of Economics*.
- Barndorff-Nielsen, O. E. 1997. Normal inverse gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*. pp. 1-13.
- Bates, D. S. 1996. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *Review of financial studies*, 9(1):69–107.
- BCBS 1988. International convergence of capital measurement and capital standards. <http://www.bis.org/publ/bcbs04a.pdf>. [Online; accessed 12-Jun-2016].
- BCBS 1996. Amendment to the capital accord to incorporate market risks. <http://www.bis.org/publ/bcbs24.pdf>. [Online; accessed 12-Jun-2016].
- BCBS 2004. Basel II: International Convergence of Capital Measurement and Capital Standards: a Revised Framework. <http://www.bis.org/publ/bcbs107.pdf>. [Online; accessed 12-Jun-2016].
- BCBS 2006. Basel II: Revised international capital framework. <http://www.bis.org/publ/bcbs128.pdf>. [Online; accessed 12-Jun-2016].
- Beem, J. 2010. Credit risk modeling and CDS valuation: an analysis of structural models.
- Bielecki, T. and Rutkowski, M. 2004. *Credit Risk: Modeling, Valuation and Hedging*. Springer, Heidelberg, Germany, 1st edition. 540 p.
- Bingham, N. H. and Kiesel, R. 2013. *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*. Springer Science & Business Media.
- Björk, T. 2009. *Arbitrage Theory in Continuous Time*. Oxford University Press, USA, 3rd edition. 512 p.
- Black, F. and Cox, J. C. 1976. Valuing corporate securities: Some effects of bond indenture provisions. *The Journal of Finance*, 31(2):351–367.
- Black, F., Derman, E., and Toy, W. 1990. A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal*, 24–32.
- Black, F. and Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.
- Brigo, D. 2010. *Credit models and the crisis a journey into CDOs, Copulas, correlations and dynamic models*. John Wiley & Sons, Chichester, West Sussex.

- Brigo, D. and Mercurio, F. 2006. *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*. Springer, Heidelberg, Germany, 2nd edition. 1037 p.
- Brigo, D., Morini, M., and Pallavicini, A. 2013. *Counterparty Credit Risk, Collateral and Funding: With Pricing Cases For All Asset Classes*. The Wiley Finance Series. Wiley.
- Cariboni, J. and Schoutens, W. 2009. Jumps in intensity models: investigating the performance of Ornstein-Uhlenbeck processes in credit risk modeling. *Metrika*, 69(2-3):173–198.
- Carr, P., Geman, H., Madan, D., and Yor, M. 2002. The fine structure of asset returns: An empirical investigation. *Journal of Business*, 75:305–332.
- Carr, P., Geman, H., Madan, D. B., and Yor, M. 2003. Stochastic volatility for Lévy processes. *Mathematical Finance*, 13(3):345–382.
- Cerny, J. and Witzany, J. 2014. Interest rate swap credit valuation adjustment. *The Journal of Derivatives*, 23(2):24–35.
- Chen, L., Collin-Dufresne, P., and Goldstein, R. S. 2009. On the relation between the credit spread puzzle and the equity premium puzzle. *Review of Financial Studies*, 22(9):3367–3409.
- Choroś-Tomczyk, B., Härdle, W. K., and Okhrin, O. 2016. A semiparametric factor model for cdo surfaces dynamics. *Journal of Multivariate Analysis*, 146:151–163.
- Choroś-Tomczyk, B., Härdle, W. K., and Overbeck, L. 2014. Copula dynamics in CDOs. *Quantitative Finance*, 14(9):1573–1585.
- Christoffersen, P., Jacobs, K., Jin, X., and Langlois, H. 2013. Dynamic dependence in corporate credit. Technical report, working paper, Bauer College of Business, University of Houston.
- Clewlow, L. and Strickland, C. 1998. *Implementing Derivative Models*. Wiley, USA, 1st edition. 330 p.
- Cont, R. and Minca, A. 2013. Recovering portfolio default intensities implied by cdo quotes. *Mathematical Finance*, 23(1):94–121.
- Cooley, J. W. and Tukey, J. W. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of computation*, 19(90):297–301.
- Cossin, D. 2001. *Advanced credit risk analysis : financial approaches and mathematical models to assess, price, and manage credit risk*. John Wiley & Sons, New York.
- Cox, J., Ingersoll, J., and Ross, S. 1985. A theory of the term structure of interest rates. *Econometrica*, 53:385–407.
- Cox, J., Ross, S., and Rubinstein, M. 1979. Option pricing: A simplified approach. *Journal of Financial Economics*, 7.

- Craig, I. J. and Sneyd, A. D. 1988. An alternating-direction implicit scheme for parabolic equations with mixed derivatives. *Computers and Mathematics with Applications*, 16(4):341–350.
- CreditSuisse 1997. *CreditRisk⁺. A credit risk management framework. (Technical document)*. Credit Suisse - First Boston. 71 p.
- Crépey, S. 2013. *Financial modeling a backward stochastic differential equations perspective*. Springer, Berlin New York.
- Crosbie, P. and Bohn, J. 2003. *Modeling default risk*. Moody's KMV working paper. 31 p.
- Crouhy, M. 2014. *The essentials of risk management*. McGraw-Hill, New York.
- Cryer, C. 2014. *A Math Primer for Engineers*. EBL-Schweitzer. IOS Press.
- Deventer, D. 2013. *Advanced financial risk management : tools and techniques for integrated credit risk and interest rate risk management*. Wiley, Singapore.
- Duffie, D. 2003. *Credit risk : pricing, measurement, and management*. Princeton University Press, Princeton, N.J.
- Duffie, D. 2010. *Dynamic asset pricing theory*. Princeton University Press.
- Duffie, D. and Singleton, K. J. 1999. Modeling term structures of defaultable bonds. *Review of Financial studies*, 12(4):687–720.
- Duffie, D. and Singleton, K. J. 2012. *Credit Risk: Pricing, Measurement, and Management*. Princeton Series in Finance. Princeton University Press.
- Duffy, D. 2006. *Finite difference methods in financial engineering : a partial differential equation approach*. John Wiley, Chichester, England Hoboken, NJ.
- Embrechts, P., Lindskog, F., and McNeil, A. 2001. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich*.
- Feldhütter, P. and Schaefer, S. 2015. The credit spread puzzle in the Merton model-myth or reality?
- Fusai, G. and Roncoroni, A. 2008. *Implementing Models in Quantitative Finance*. Springer, USA, 1st edition. 608 p.
- Garman, M. B. and Kohlhagen, S. W. 1983. Foreign currency option values. *Journal of international Money and Finance*, 2(3):231–237.
- Gatheral, J. 2006. *The volatility surface : a practitioner's guide*. John Wiley & Sons, Hoboken, N.J.
- Gatheral, J. and Jacquier, A. 2014. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71.

- Geman, H., El Karoui, N., and Rochet, J.-C. 1995. Changes of numeraire, changes of probability measure and option pricing. *Journal of Applied probability*, pages 443–458.
- Geske, R. 1979. The valuation of compound options. *Journal of financial economics*, 7(1):63–81.
- Gikhman, I. I. 1972. *Stochastic differential equations*. Springer-Verlag, Berlin, New York.
- Gil-Pelaez, J. 1951. Note on the inversion theorem. *Biometrika*, 38.
- Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. Springer, USA, 1st edition. 596 p.
- Gordy, M. B. 2003. A risk-factor model foundation for ratings-based bank capital rules. *Journal of financial intermediation*, 12(3):199–232.
- Gregory, J. 2012. *Counterparty Credit Risk and Credit Value Adjustment: A Continuing Challenge for Global Financial Markets*. Wiley, New York, USA. 480 p.
- Gregory, J., Laurent, J., and Burtschell, X. 2007. Beyond the gaussian copula: Stochastic and local correlation. 27 p.
- Gupton, G., Finger, C., and Bhatia, M. 1997. *CreditMetrics – Technical Document*. J.P. Morgan, New York, USA. 212 p.
- Harrison, J. M. and Kreps, D. M. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic theory*, 20(3):381–408.
- Haworth, H., Reisinger, C., and Shaw, W. 2008. Modelling bonds and credit default swaps using a structural model with contagion. *Quantitative Finance*, 8(7):669–680.
- Heath, D., Jarrow, R., and Morton, A. 1990. Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation. *Journal of Financial and Quantitative Analysis*, 25.
- Heath, D., Jarrow, R., and Morton, A. 1991. Contingent Claims Valuation with a Random Evolution of Interest Rates. *Review of Futures Markets*, 9.
- Heath, D., Jarrow, R., and Morton, A. 1992. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica*, 60(1).
- Heston, S. L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343.
- Hippler, S. 2008. *Pricing bermudan swaptions in the LIBOR market model*. PhD thesis, University of Oxford.
- Hirsa, A. 2016. *Computational Methods in Finance*. Chapman and Hall/CRC Financial Mathematics Series. CRC Press.

- Hirsa, A. and Neftci, S. 2013. *An Introduction to the Mathematics of Financial Derivatives*. Elsevier Science.
- Ho, T. and Lee, S. 1986. Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, 41:1011–1029.
- Hoffman, J. and Frankel, S. 2001. *Numerical Methods for Engineers and Scientists, Second Edition*,. Taylor & Francis.
- Hogan, M. and Weintraub, K. 1993. The lognormal interest rate model and Eurodollar futures. *Citibank, New York*.
- Huang, J.-Z. and Huang, M. 2012. How much of the corporate-treasury yield spread is due to credit risk? *Review of Asset Pricing Studies*, 2(2):153–202.
- Hull, J. 2008. *Options, Futures, and Other Derivatives*. Prentice Hall, USA, 7th edition. 848 p.
- Hull, J., Nelken, I., and White, A. 2004. Merton’s model, credit risk, and volatility skews. *Journal of Credit Risk Volume*, 1(1):05.
- Hull, J. and White, A. 1990. Pricing interest-rate derivative securities. *The Review of Financial Studies*, 3:573–592.
- Hull, J. and White, A. 2004. Valuation of a CDO and an n-th to Default CDS Without Monte Carlo Simulation. *Journal of Derivatives*, 12.
- Hull, J. C. and White, A. 2001. Valuing credit default swaps II: modeling default correlations. *Journal of Derivatives*, 8(3):12–22.
- Hull, J. C. and White, A. D. 2003. The valuation of credit default swap options. *The journal of derivatives*, 10(3):40–50.
- Hurst, S. 1997. *On the Stochastic Dynamics of Stock Market Volatility*. Australian National University.
- Jarrow, R. and Protter, P. 2004. Structural versus reduced form models: a new information based perspective. *Journal of Investment management*, 2(2):1–10.
- Jarrow, R. and Turnbull, S. 1995. Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50:53–85.
- Jarrow, R. A., Lando, D., and Turnbull, S. M. 1997. A Markov model for the term structure of credit risk spreads. *Review of Financial studies*, 10(2):481–523.
- Jones, S. 2009. Of couples and copulas: the formula that felled Wall St. *Financial Times*.
- Kalemanova, A., Schmidt, B., and Werner, R. 2005. The Normal inverse Gaussian distribution for synthetic CDO pricing. 20 p.
- Kane, D. O. 2008. *Modelling Single-name and Multi-name Credit Derivatives*. Wiley, USA, 1st edition. 514 p.

- Kenyon, C. and Stamm, R. 2012. *Discounting, LIBOR, CVA and Funding: Interest Rate and Credit Pricing*. Applied Quantitative Finance. Palgrave Macmillan UK.
- Kolman, M. 2014a. A one-factor copula-based model for credit portfolios. *Journal of Risk*, 17(2).
- Kolman, M. 2014b. Option pricing using Fourier-Cosine expansion. *Risk Management*.
- Kolman, M. 2015a. A defaultable binomial tree. *CEREI Journal*, (under review process).
- Kolman, M. 2015b. Numerical solutions to higher-order PDEs with applications to finance. *Advanced methods of quantitative finance*.
- Kolman, M. 2016. Pricing credit using PDE techniques. *Advanced Methods of Computational Finance*, (awaiting publication).
- Kolman, M., Privara, S., and Witzany, J. 2014. Recovery rates in consumer lending: Empirical evidence and the model comparison. *Bulletin of the Czech Econometric Society*, 21(32).
- Kou, S. G. 2002. A jump-diffusion model for option pricing. *Management science*, 48(8):1086–1101.
- Kwok, Y. K. 2008. *Mathematical models of financial derivatives*. Springer, Berlin.
- Lall, R. 2012. From failure to failure: The politics of international banking regulation. *Review of International Political Economy*, 19(4):609–638.
- Lando, D. 1998. On Cox processes and credit risky securities. *Review of Derivatives Research*, 2:99–120.
- Lee McGinty, Rishad Ahluwalia, M. W. 2004. A model for base correlation calculation. *Working paper*.
- Leland, H. E. 1994. Corporate debt value, bond covenants, and optimal capital structure. *The journal of finance*, 49(4):1213–1252.
- Li, D. X. 1999. On default correlation: A copula function approach. *Available at SSRN 187289*.
- Lichters, R., Stamm, R., and Gallagher, D. 2015. *Modern Derivatives Pricing and Credit Exposure Analysis: Theory and Practice of CSA and XVA Pricing, Exposure Simulation and Backtesting*. Applied Quantitative Finance. Palgrave Macmillan UK.
- Loffler, G. 2007. *Credit risk modeling using Excel and VBA*. Wiley, Chichester, England Hoboken, NJ.
- Longin, F. and Solnik, B. 2001. Extreme correlation of international equity markets. *The journal of finance*, 56(2):649–676.
- Loretan, M. and English, W. B. 2000. Evaluating correlation breakdowns during periods

- of market volatility. *Board of Governors of the Federal Reserve System International Finance Working Paper*, (658).
- Madan, D., Carr, P., and Chang, E. 1998. The variance gamma process and option pricing. *European Finance Review*, 2:79–105.
- Madan, D. and Seneta, E. 1990. The variance gamma model for share market returns. *The Journal of Business*, 63:511–524.
- Málek, J. 2005. *Dynamika úrokových měř a úrokové deriváty*. Ekopress, Praha.
- Margrabe, W. 1978. The value of an option to exchange one asset for another. *The journal of finance*, 33(1):177–186.
- Marshall, A. W. and Olkin, I. 1988. Families of multivariate distributions. *Journal of the American Statistical Association*, 83(403):834–841.
- McNeil, A. J., Frey, R., and Embrechts, P. 2005. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press.
- Merton, R. C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, 29(2):449–470.
- Merton, R. C. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of financial economics*, 3(1-2):125–144.
- Miu, P. and Ozdemir, B. 2006. Basel requirement of downturn LGD: Modeling and estimating PD and LGD correlations. *Journal of Credit Risk*, 2(2):43–68.
- Musiela, M. and Rutkowski, M. 2010. *Martingale Methods in Financial Modelling*. Springer, Heidelberg, Germany, 1st edition. 638 p.
- Okhrin, O., Okhrin, Y., and Schmid, W. 2013. On the structure and estimation of hierarchical archimedean copulas. *Journal of Econometrics*, 173(2):189–204.
- Ou, S., Chiu, D., Wen, B., and Metz, A. 2016. Annual default study: Corporate default and recovery rates, 1920-2015. *Special Comment. New York: Moody's Investors Services*, pages 1–76.
- Pikulin, V. P. 2001. *Equations in mathematical physics : a practical course*. Birkhäuser, Basel Boston.
- Protter, P. E. 2005. *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Pykhtin, M. 2004. Portfolio credit risk multi-factor adjustment. *Risk Magazine*, 17(3):85–90.
- Rachev, S. 2004. *Handbook of Computational and Numerical Methods in Finance*. Birkhäuser Boston Imprint Birkhäuser, Boston, MA.

- Radová, J., Dvořák, P., and Málek, J. 2009. *Finanční matematika pro každého - 7. aktualizované vydání*. Finance (Grada). Grada.
- Rajan, A. 2007. *The structured credit handbook*. Wiley, Hoboken, NJ.
- Rouah, F. D. 2013. *The Heston model and its extensions in MATLAB and C#*. John Wiley and Sons.
- Ruppert, D. 2011. *Statistics and data analysis for financial engineering*. Springer, New York.
- Salmon, F. 2012. The formula that killed Wall Street. *Significance*, 9(1):16–20.
- Sandoval, L. and Franca, I. D. P. 2012. Correlation of financial markets in times of crisis. *Physica A: Statistical Mechanics and its Applications*, 391(1):187–208.
- Schlosser, A. 2011. *Pricing and Risk Management of Synthetic CDOs*. Springer, USA, 1st edition. 282 p.
- Schmid, B. 2004. *Credit risk pricing models : theory and practice*. Springer, Berlin New York.
- Schönbucher, P. 2003. *Credit Derivatives Pricing Models: Models, Pricing and Implementation*. Wiley, USA, 1st edition. 396 p.
- Schönbucher, P. J. and Schubert, D. 2001. Copula-dependent default risk in intensity models.
- Servigny, A. and Jobst, N. 2007. *The handbook of structured finance*. McGraw-Hill, New York.
- Shephard, N. G. 1991. From characteristic function to distribution function: a simple framework for the theory. *Econometric theory*, 7(04):519–529.
- Shimko, D. C., Tejima, N., and Van Deventer, D. R. 1993. The pricing of risky debt when interest rates are stochastic. *The Journal of Fixed Income*, 3(2):58–65.
- Shreve, S. E. 2004. *Stochastic calculus for finance II: Continuous-time models*, volume 11. Springer.
- Synowiec, D. 2008. Jump-diffusion models with constant parameters for financial log-return processes. *Computers and Mathematics with Applications*, 56(8):2120–2127.
- Tapiero, C. 2013. *Engineering risk and finance*. Springer, New York.
- Tavella, D. and Randall, C. 2000. *Finite Differences in Pricing Financial Instruments*. Wiley.
- Ugail, H. 2011. *Partial Differential Equations for Geometric Design*. SpringerLink : Bücher. Springer London.
- Vasicek, O. 1977. An Equilibrium Characterisation of the Term Structure. *Journal of Financial Economics*, 5:177–188.

- Vasicek, O. 1991. Limiting loan loss probability distribution. *Finance, Economics and Mathematics*, pages 147–148.
- Vasicek, O. 2002. The Distribution of Loan Portfolio Value. *Risk Magazine*. pp. 160-162.
- Vrins, F. D. 2009. Double t copula pricing of structured credit products: Practical aspects of a trustworthy implementation. *Journal of Credit Risk*.
- Wilmott, P., Howison, S., and Dewynne, J. 1995. *The mathematics of financial derivatives: a student introduction*. Cambridge University Press.
- Witzany, J. 2009. Estimating LGD Correlation. *Institute of Economic Studies, Faculty of Social Sciences. Prague*. working paper series.
- Witzany, J. 2013. A note on the Vasicek’s model with the logistic distribution. *Ekonomický časopis*, (10):1053–1066.
- Yu, F. 2007. Correlated defaults in intensity-based models. *Mathematical Finance*, 17(2):155–173.
- Zhang, B. Y., Zhou, H., and Zhu, H. 2009. Explaining credit default swap spreads with the equity volatility and jump risks of individual firms. *Review of Financial Studies*, 22(12):5099–5131.
- Zhang, J. and Thomas, L. C. 2012. Comparisons of Linear Regression and Survival Analysis using Single and Mixture Distributions Approaches in Modelling LGD. *International Journal of Forecasting*, 28.
- Zhou, C. 1997. A jump-diffusion approach to modeling credit risk and valuing defaultable securities. *Available at SSRN 39800*.
- Zikanov, O. 2010. *Essential computational fluid dynamics*. Wiley, Hoboken, N.J.