A thesis submitted in partial fulfilment of the requirements for the degree of Candidatus Mercaturae in Mathematics: cand.merc.(mat.)

# Credit Value Adjustment

## Pricing Wrong Way Risk on Interest Rate Swaps


*Author:*

Magnus Mencke

*Student number:*

109766

*Supervisor:*

David Glavind Skovmand


*Number of pages:*

72 pages (142 573 characters)


Department of Finance
Copenhagen Business School


17th May 2021

# Kreditværdijustering

Værdiansættelse af *wrong way risk* på renteswaps.


## Resumé

Denne kandidatafhandling omhandler værdiansættelse af kreditrisiko på renteswaps. Først forklares arbitrage-teori generelt, dernæst bliver teoretiske priser for rentederivater, kreditderivater og modpartsrisiko udledt og slutteligt bliver den bilaterale kreditværdijustering (BVA) beregnet ved hjælp af Monte Carlo-simulering ved tre forskellige korrelationsstruktur mellem renter og fallitintensiteter; uafhængighed, *wrong way risk* og *right way risk*.

Det vises, at når parterne kan gå konkurs, så er værdiansættelse af swaps ikke længere modeluafhængig, men afhænger af den fælles udvikling i renter og fallitintensiteter. Renten er modelleret ved en to-faktor Gaussisk model for den korte rente betegnet G2++, hvilket er et specialtilfælde af Heath-Jarrow-Morton modelrammen. Fallitintensiteten er modelleret ved en udvidet Cox-Ingersoll-Ross model betegnet CIR++. Modellen for den korte rente er kalibreret til swap- og swaptionspriser og fallitintensitetsmodellen er kalibreret til kreditspænd på CDS-kontrakter. Modellerne kan reproducere priserne på lineære produkter nøjagtigt, mens de kun reproducerer volatilitetsoverfladen i nogen grad. Modellerne er således konsistente med nylig markedsdata, hvilket er en forudsætning for enhver fornuftig finansiel model. Under antagelse om uafhængighed og udsat konkurs kan BVA approksimeres som en portefølge af swaptioner, hvor porteføljevægtene afhænger af sandsynligheden for konkurs.

*Wrong-* og *right way risk* indføres i modellen ved at antage korrelation mellem Wiener-processerne, der henholdsvis driver renten og modpartens fallitintensitet.

# Contents

# 1

# Introduction

## 1.1 Background

Research on counterparty credit risk dates at least back to the 1990s with [Sorensen and Bollier (1994)] as an early example of the application to interest rate swaps. However, the research area did not attract much attention before the Global Financial Crisis of 2007-2008. Before the crisis credit risk was typically seen as negligible as the Federal Reserve was expected to support large banks in situations of financial distress. Informally this is called the 'too big to fail' theory. With the default of Lehman Brothers on 15$^{\text{th}}$ September 2008 it became clear that the typical assumption in derivatives pricing of a default-free counterparty did not hold in practice. Lehman Brothers was a huge player in the *over-the-counter* (OTC) derivatives market and before bankruptcy they had an estimated 5% market share globally, which is equivalent of $35 trillion notional. At the time of bankruptcy they had over 900 000 derivatives transactions and claims of $1.2 trillion was made towards the estate [Fleming and Sarkar (2014)]. This made it clear that the credit risk in derivatives transactions needs to be accounted for and today it is part of accounting standards on valuation of derivatives (IFRS 13). Furthermore regulation of OTC derivatives have been increased significantly since the crisis through 'Dodd-Frank Act' in the United States of America and 'EMIR' in the European Union.

The reason for focusing on interest rate swaps in this thesis is twofold: i) As the value of an interest rate swap can be both positive and negative both parties are exposed to credit risk in contrast to options for example, and ii) the market for interest rate derivatives are by far the largest class of derivatives.

Documentation from the International Swaps and Derivatives Association (ISDA) will occasionally be referred to in this thesis. ISDA is an association working with standardising of financial derivatives transactions. An example is the ISDA Master Agreement, which is the standard document to govern OTC derivatives transactions. ISDA also provides statistics about transactions. Through the first quarter of 2021 348 000 fixed-for-floating interest rate swaps trades were made amounting to a notional of $27.6 trillion globally. To put the number into relief the GDP of the United States of America was $21.5 trillion in 2020[1], so the market for interest rate swaps is very large. 75.1% of the traded notional was with central clearing and interest rate swaps in Euro account for $4.0 trillion traded notional through the quarter [ISDA (2021)]. As of 31$^{\text{st}}$ December 2020 the total notional outstanding was 355.8 trillion[2].

## 1.2 Research Question

From the case of Lehman Brothers it is seen that credit risk can have a big impact on the value of derivatives and from the statistics it is seen that interest rate swaps are widely used in practice. The focus of this thesis is on credit risk in relation to interest rate swaps and in particular how credit risk affects the pricing. This motivates the research question

> **How can the Credit Value Adjustment on an Interest Rate Swap be determined under Wrong Way Risk?**

[1] Source: `https://fred.stlouisfed.org/series/GDP`.

[2] Source: `http://swapsinfo.org/derivatives-notional-outstanding/`.

The research question involves a couple definitions that has not been made yet. Credit Value Adjustment is defined in the beginning of section 4.3 and Wrong Way Risk is defined in definition 4.7. To avoid repetition the definitions are not listed here as well.

### 1.2.1 Delimitation

The theory presented in this thesis is primarily relevant for uncollateralised swaps without central clearing, but some of the theory is also applicable in special cases of collateralised swaps and for swaps with central clearing.

The interest rate models that have been investigated have been limited to the special case of Heath-Jarrow-Morton (HJM) models, where the short rate is both Gaussian and Markovian. This is important for tractability of the models, which both ensures that the models are easily interpreted and computationally efficient. Other interesting HJM models are *Libor Market Model* and *Quasi-Gaussian* models, but these are significantly more complex to understand and implement. Cash flows are discounted on the same curve used to forecast the floating rate on the swap, so it is assumed that the bank can fund itself at this rate.

In terms of credit risk modelling a reduced-form approach has been taken. This is chosen as it is far easier to calibrate an intensity model to market data compared to calibration in a structural model for credit risk. The reduced-form approach is also rather flexible making it possible to eventually extending the model.

Monte Carlo is the numerical method of choice as it is flexible and easy to implement. Finite Difference methods are not really applicable to CVA as it does not work for path dependent derivatives, which CVA is. Correlation between interest rates and intensities are introduced through correlation of the Wiener processes that drives interest rates and intensities. A more sophisticated approach could have been to model the dependence trough a copula.

## 1.3 Scientific Method

The scientific method used in this thesis is the axiomatic deductive method. A mathematical approach to understanding financial markets has been chosen, but it has been important to make as few simplifying assumptions as possible. This philosophy can be encapsulated by the following quote

> *I will remember that I didn't make the world, and it doesn't satisfy my equations.*
>
> *Though I will use models boldly to estimate value, I will not be overly impressed by mathematics.*
>
> *I will never sacrifice reality for elegance without explaining why I have done so.*
>
> *Nor will I give the people who use my model false comfort about its accuracy. Instead, I will make explicit its assumptions and oversights.*
>
> (Derman and Wilmott 2009)

Even though the approach of this thesis is first and foremost theoretical, it has also been a goal to produce sound numbers making it possible to use the theory in practice. Sometimes intuition will be relied upon rather than mathematical rigor and it is almost always assessed whether the mathematical expressions also make economic sense. In terms of language, the pronoun 'we' will be used referring to 'the author and the reader'.

The thesis is largely based on [Andersen and Piterbarg (2010)] and [Brigo, Morini and Pallavicini (2013)].

## 1.4 Structure of Thesis

The rest of the thesis is structured as follows:

- Chapter 2 describes general arbitrage theory and how it can be used to price derivatives – both analytical and by means of numerical methods.
- Chapter 3 describes the modelling framework for pricing fixed income derivatives.
- Chapter 4 describes the modelling framework for pricing credit derivatives and determining credit risk on derivatives transactions in general.
- Chapter 5 consists of application of the theory through implementation of the models in `QuantLib` and a case study of a concrete use case for interest rate swaps.
- Lastly the thesis is concluded and further research is suggested in the perspective.

# 2

# Arbitrage Pricing Theory

The martingale approach to arbitrage theory is described in detail through this chapter. The chapter is primarily based on chapters 1 and 3 in [Andersen and Piterbarg (2010)]. Only the concepts that will be used in this thesis are explained here.

## 2.1 Stochastic Processes

Suppose we have a stochastic process $X = \{X(t) : 0 \leq t < \infty\}$ on the probability space $(\Omega, \mathscr{F}, \mathbf{P})$. $\mathbf{P}$ is a probability measure on the sample space $(\Omega, \mathscr{F})$. A sample path (trajectory) for $X$ is defined by the function $t \mapsto X(t, \omega)$ for a fixed sample point $\omega \in \Omega$. The expectation operator under probability measure $\mathbf{P}$ will be denoted by $\mathbb{E}^{\mathbf{P}}[\,\cdot\,]$ and in a similar fashion the probability operator will be denoted by $\mathbb{P}^{\mathbf{P}}[\,\cdot\,]$. As we are working with a lot of different probability measures, this is a simple and consistent notation[3].

The filtration $\mathscr{F} = \{\mathscr{F}_t\}_{t \geq 0}$ is non-decreasing with time: $\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F}$ for $0 \leq s \leq t < \infty$. The simplest choice of filtration is the natural filtration of the stochastic process $X$: $\mathscr{F}_t = \sigma(X(s) : 0 \leq s \leq t)$, which can be thought of as the information generated by the process up until the time $t$. Formally this is the smallest $\sigma$-algebra on $\Omega$ such that $X(t)$ is measurable for any $t$. This choice of filtration will obviously make $X(t)$ a $\mathscr{F}_t$-measurable random variable for all $t$ and $X$ is thus said to be adapted to the filtration. Typically we will only be concerned about the stochastic process within finite time, e.g. $t \in [0, T]$ where $T$ is the final time of interest. We have the usual technical conditions that the filtration must be right-continuous such that $\cap_{s \geq t} \mathscr{F}_s = \mathscr{F}_t$ for all $t$ and complete such that $\mathscr{F}_0$ contains all $\mathbf{P}$ null sets. This means that there cannot be any jumps in information over a small time interval and that $\mathscr{F}_0$ contains all subsets of zero $\mathbf{P}$-probability.

**Proposition 2.1 (Tower Property).** *Consider the times $s < t < T$. Assume that $X$ is adapted to $\mathscr{F}$ and that $\mathbb{E}[|X(t)|] < \infty$ for all $t$. We then have*

$$\mathbb{E}[\mathbb{E}[X(T)|\mathscr{F}_t]|\mathscr{F}_s] = \mathbb{E}[X(T)|\mathscr{F}_s] \quad a.s. \tag{2.1}$$

The abbreviation a.s. stands for *almost surely*, which means that an event happens with probability one. In general explicit expressions for expectations hold almost surely, but this will not always be stated explicitly. The tower property is sometimes also called the law of iterated expectations or the law of total expectation. This result has the interpretation that it does not gives us any more information today to condition on future information.

[3]Usually the probability is denoted by $\mathbb{P}$, such that the expectation operator becomes $\mathbb{E}^{\mathbb{P}}[\,\cdot\,]$ and the probability measure and operator coincides: $\mathbb{P}[\,\cdot\,]$. When for example working with the $T_i$-forward measure this will become $\mathbb{T}_i$ which makes the notation for similar expressions very different – this is avoided by choosing the rather unorthodox (but self-explanatory) notation described here.

*Proof.* We have that $\mathscr{F}_s \subseteq \mathscr{F}_t$, so $\mathscr{F}_s \cap \mathscr{F}_t = \mathscr{F}_s$. This gives us

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[X(T)|\mathscr{F}_t\right]|\mathscr{F}_s\right] &= \int_{A\in\mathscr{F}_s} \mathbb{E}\left[X(T)|\mathscr{F}_t\right] \mathrm{d}\mathbf{P}(A) \\ &= \int_{A\in\mathscr{F}_s} X(T)\, \mathrm{d}\mathbf{P}(A) \\ &= \mathbb{E}\left[X(T)|\mathscr{F}_s\right]\end{aligned}$$

■

**Definition 2.2 (Markov Property).** The stochastic process $X(t)$ is said to be a Markov process if the following equality hold for times $0 \leq t < T$

$$\mathbb{E}^{\mathbf{P}}\left[X(T)|\mathscr{F}_t\right] = \mathbb{E}^{\mathbf{P}}\left[X(T)|X(t)\right] \tag{2.2}$$

The Markov property is useful as it means that we only have to know the current value of $X$ (and not the entire path) to determine the expectation.

**Definition 2.3 (Stopping Time).** A random time $\tau$ is a *stopping time* of the filtration, if the event $\{\tau \leq t\}$ belongs to the filtration $\mathscr{F}_t$.

We will typically think of the random time $\tau$ as the time of default. Based on the information $\mathscr{F}_t$ we know whether or not the default has occurred yet.

**Definition 2.4 (Martingales).** We still have times $s < t$. Assume that $\mathbb{E}\left[|X_t|\right] < \infty$ for all $t \geq 0$ and that $X$ is adapted to $\mathscr{F}$. The process $X$ is a submartingale if we have $\mathbb{E}\left[X_t|\mathscr{F}_s\right] \geq X_s$ and is a supermartingale if we have $\mathbb{E}\left[X_t|\mathscr{F}_s\right] \leq X_s$. The process is a martingale if it is both a sub- and a supermartingale, i.e. $\mathbb{E}\left[X_t|\mathscr{F}_s\right] = X_s$.

The usefulness of the martingale property in relation to arbitrage pricing will be shown later on. In relation to martingales, the martingale representation theorem is important. It means that we are able to define local martingales in terms of a driftless version of the SDE given in equation 2.4.

**Definition 2.5 (Wiener Process).** A Wiener process $W$ is a continuous adapted process defined on some probability space $(\Omega, \mathscr{F}, \mathbf{P})$ with the properties

1. $W(0) = 0$ a.s.
2. For $0 \leq s < t$ the increment $W(t) - W(s)$ is independent of $\mathscr{F}_s$ and normally distributed with mean zero and variance $t - s$.

Using a stochastic integral the Wiener process can be written as

$$W(t) = \int_0^t \mathrm{d}W(s) \tag{2.3}$$

and it can be used to construct an Itô-process $X$

$$\mathrm{d}X(s) = \mu(t, X(s))\,\mathrm{d}s + \sigma(t, X(s))\,\mathrm{d}W(s) \tag{2.4}$$

where $X$ is defined in terms of its *stochastic differential equation* (SDE). By integrating the expression we obtain the integral equation

$$X(t) = X(0) + \int_0^t \mu(s, X(s))\,\mathrm{d}s + \int_0^t \sigma(s, X(s))\,\mathrm{d}W(s) \tag{2.5}$$

An Itô process is also referred to as a diffusion process, where $\mu(t, X(s))$ is the drift-parameter and $\sigma(s, X(s))$ is the diffusion-parameter. The Itô integral $\int_0^t \sigma(s, X(s))\,\mathrm{d}W(s)$ is an $\mathscr{F}_t$-measurable martingale with mean zero. We have the regularity conditions

$$\int_0^t |\mu(s, X(s))|\,\mathrm{d}s < \infty \qquad \text{for all } t \quad \text{a.s.} \tag{2.6}$$

and

$$\int_0^t \sigma^2(s, X(s))\,\mathrm{d}s < \infty \qquad \text{for all } t \quad \text{a.s.} \tag{2.7}$$

The main result of Itô calculus is Itô's lemma

**Proposition 2.6 (Itô's Lemma).** *Assume that $f(t, x)$ is in $C^2$ such that all relevant partial derivatives exist. Assume that $X(t)$ is an Itô process and define $Y(t) := f(t, X(t))$. The dynamics of $Y(t)$ is then given by*

$$\mathrm{d}Y(t) = f_t\,\mathrm{d}t + f_x\,\mathrm{d}X(t) + \frac{1}{2} f_{xx} (\mathrm{d}X(t))^2 \tag{2.8}$$

*with the multiplication rules*

$$(\mathrm{d}t)^2 = 0, \qquad (\mathrm{d}t)(\mathrm{d}W(t)) = 0, \qquad (\mathrm{d}W(t))^2 = \mathrm{d}t$$

A proof of the lemma is beyond the scope of this thesis, but a heuristic argument is stated below. The reader is referred to [Karatzas and Shreve (1988)] for a formal proof.

*Heuristic argument.* We can write the (approximate) dynamics of the stochastic process $Y(t) = f(t, X(t))$ using a second-order Taylor expansion

$$\mathrm{d}Y(t) = f_t\,\mathrm{d}t + f_x\,\mathrm{d}X(t) + \frac{1}{2}\left(f_{tt}(\mathrm{d}t)^2 + f_{xx}(\mathrm{d}X(t))^2 + 2 f_{tx}\,\mathrm{d}t\,\mathrm{d}X(t)\right)$$

Subscript denote the partial derivative with respect to that variable. Assume that $(\mathrm{d}t)^2 = 0$, $\mathrm{d}t\,\mathrm{d}W(t) = 0$ and $(\mathrm{d}W(t))^2 = \mathrm{d}t$. This is a reasonable assumption for small $\mathrm{d}t$ and true in the limit $\mathrm{d}t \downarrow 0$, i.e. the size of the increments going to zero from above. This assumption makes the second-order Taylor approximation exact as all third-order elements would then be zero. We can then write

$$\mathrm{d}Y(t) = f_t\,\mathrm{d}t + f_x\,\mathrm{d}X(t) + \frac{1}{2} f_{xx} (\mathrm{d}X(t))^2 \tag{2.9}$$

We have now arrived at Itô's lemma put forward in Proposition 4.11 in [Björk (2009), p. 52]. This is easily extendable to the case where we have an $n$-dimensional stochastic process $X$ driven by $n$ Wiener processes. Assume that the correlation between two Wiener processes, $W_i(t)$ and $W_j(t)$, is given by $\rho_{ij}$ such that $\mathrm{d}W_i(t)\,\mathrm{d}W_j(t) = \rho_{ij}\,\mathrm{d}t$. We then have the Itô formula

$$\mathrm{d}Y(t) = f_t\,\mathrm{d}t + f_x^\top \cdot \mathrm{d}X(t) + \frac{1}{2}\,\mathrm{d}X(t)^\top \cdot f_{xx} \cdot \mathrm{d}X(t) \tag{2.10}$$

where $f_x$ is now the gradient and $f_{xx}$ is now the Hessian matrix. We have the regularity conditions that the drift vector must be in $L^1$ and the squared diffusion vector must be in $L^2$ [Andersen and Piterbarg (2010), p. 4]. This is an extension of equations (2.6) and (2.7) to the vector case. It should be noted that the correlation term does not appear explicitly

## 2.2 Probability Theory and Arbitrage Opportunities

The no-arbitrage principle is equivalent to the common saying 'there is no such thing as a free lunch'. It is the notion that you cannot get something from nothing and the formal definition is given by

**Definition 2.7 (Arbitrage).** An arbitrage opportunity is an initially costless, self-financing strategy that for some $t$ always has a non-negative payoff and have a positive probability for a strictly positive payoff.

For the model to be realistic it has to be free of arbitrage as rational arbitrageurs would exploit the mispricing.

The two probability measures, $\mathbf{P}$ and $\mathbf{Q}$, on the common measure space $(\Omega, \mathscr{F})$ are equivalent if $\mathbf{P}(A) = 0 \iff \mathbf{Q}(A) = 0, \forall\, A \in \mathscr{F}$. If this is the case we can write

$$\mathbf{Q}(A) = \mathbb{E}^{\mathbf{P}}\left[\mathbb{1}_{\{A\}}\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\right], \quad \forall\, A \in \mathscr{F} \tag{2.11}$$

where $\mathbb{1}_{\{A\}}$ is the indicator function that returns 1 if $A$ is true and 0 if not. $\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}$ is the Radon-Nikodym derivative. We can associate the probability measure $\mathbf{Q}$ with a density process

$$\varsigma(t) = \mathbb{E}^{\mathbf{P}}\left[\frac{\mathrm{d}\mathbf{Q}}{\mathrm{d}\mathbf{P}}\middle|\mathscr{F}_t\right], \quad \forall\, t \geq 0 \tag{2.12}$$

We have $\varsigma(0) = 1$ due to the equivalence of the probability measures and $\varsigma(t) = \mathbb{E}^{\mathbf{P}}\left[\varsigma(T)|\mathscr{F}_t\right]$ due to the tower property, so the process is a martingale. The density process can, loose speaking, be used to 'translate' probabilities under one measure to probabilities under another

$$\mathbb{E}^{\mathbf{Q}}\left[X(T)|\mathscr{F}_t\right] = \mathbb{E}^{\mathbf{P}}\left[X(T)\frac{\varsigma(T)}{\varsigma(t)}\middle|\mathscr{F}_t\right] \tag{2.13}$$

The two main results of the martingale approach to arbitrage pricing are listed below

**Proposition 2.8 (First Fundamental Theorem of Derivatives Pricing).** *The market is essentially arbitrage-free iff there exists a strictly positive numeraire such that the discounted asset price process is a martingale under the equivalent martingale measure induced by the numeraire.*

**Proposition 2.9 (Second Fundamental Theorem of Derivatives Pricing).** *The market is complete if proposition 2.8 holds and if the numeraire induces a unique martingale measure.*

The choice of numeraire will typically be changed arbitrarily to simplify evaluation of expectations by using the martingale property. The possibility of changing measure follows from the Radon-Nikodym theorem. The typically numeraire is the bank account with associated martingale measure $\mathbf{Q}$. The economic intuition for this choice of numeraire is that it makes sense to discount cash flow with the risk-free rate. s

**Proposition 2.10 (General Pricing Formula).** *Suppose that* $\mathbf{D}$ *is the equivalent martingale measure induced by numeraire* $D(t)$*. Then* $\frac{V(t)}{D(t)}$ *is a martingale, so the value of a* $T$*-claim* $V(t)$ *is given by*

$$V(t) = D(t)\mathbb{E}^{\mathbf{D}}\left[\frac{V(T)}{D(T)}\middle|\mathscr{F}_t\right] \tag{2.14}$$

The behaviour of Wiener processes under different probability measures are linked through Girsanov's theorem.

**Proposition 2.11 (Girsanov's Theorem).** *Suppose that we want to change measure from* $\mathbf{P}$ *to* $\mathbf{Q}$*. Then the dynamics of the Wiener process under* $\mathbf{Q}$ *is given by*

$$\mathrm{d}W^{\mathbf{Q}} = \mathrm{d}W^{\mathbf{P}} - \psi\,\mathrm{d}t \tag{2.15}$$

*where* $\psi$ *is Girsanov kernel.* $-\psi$ *is the market price of risk.*

The results above is the foundation for derivatives pricing and will be made more concrete in later chapters.

## 2.3 Monte Carlo Simulation

One way to price a claim is through simulation. The idea behind Monte Carlo simulation is closely related to the result of the central limit theorem

Typically we are interested in evaluating the expectation to a stochastic variable; let us denote this $Y$. Suppose that we are able to draw a sequence of independent identically distributed (iid.) random variables of the same distribution as $Y$. Using the law of large numbers and the central limit theorem we can then say something about both the convergence of the sample mean towards the true mean as well as how wrong we are on average, i.e. the standard error. Define the sample mean

$$\bar{Y}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i \tag{2.16}$$

The law of large numbers then tells us that $\lim_{n\to\infty}\bar{Y}_n = \mu$ where $\mu$ is the true mean. This result shows that we are able to evaluate the expectation using simulation, but it does not say anything about how large $n$ should be. Define the standard error

$$s_n := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}_n\right)^2} \tag{2.17}$$

The central limit theorem then says that

$$\frac{\bar{Y}_n - \mu}{\frac{s_n}{\sqrt{n}}} \xrightarrow{d.} N(0,1) \quad \text{for } n \to \infty \tag{2.18}$$

where $\xrightarrow{d.}$ means that it converges in distribution. With confidence level $\alpha$ a normally distributed variable is between $\Phi^{-1}\left(\frac{\alpha}{2}\right) = -\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$ and $\Phi^{-1}\left(1-\frac{\alpha}{2}\right)$. This means that the true mean $\mu$ lies in the interval $\bar{Y}_n \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{s_n}{\sqrt{n}}$. It is very useful that we both know that the simulations converge and that we have an estimate for how uncertain the values are. It is generally considered good practice to always provide the confidence interval together with values generated by Monte Carlo simulation to know how precise the values are as this might be difficult to assess judging from the number of simulations alone.

### 2.3.1 Pseudo-Random Numbers

A computer cannot by itself generate true random numbers, but the second best is pseudo-random number generators[4]. We will focus on the *linear congruential generator* where the numbers follow the recurrence relation

$$I_{n+1} = (aI_n + c) \mod m \tag{2.19}$$

where $a$ is the multiplier, $c$ is the increment and $m$ is the modulus operator. As is seen the value is determinstic given starting point $I_0$. We want the period of the generator to be large meaning that it will take a long time before the sequence of numbers are repeated. We will in general use the Mersenne Twister, which in its most common form has a period of $2^{19937} - 1$, which is a Mersenne prime. The 32-bit version generates approximately uniform integers between 0 and $2^{32} - 1$, which can easily be transformed into continuous uniform variables between 0 and 1 by dividing by the upper limit. We can then use the inverse transform method to transform this uniform variable into a variable of any distribution that we want. If we denote the uniform variable $U$ and want to generate a random variable with inverse cumulative distribution (quantile) function given by $F^{-1}(u)$, then we can generate a new random variable

$$Z = F^{-1}(U) \tag{2.20}$$

that follows the desired distribution. For this to be feasible it require a fast implementation of the quantile function. With the case of a Gaussian distribution there is a fast approximation which is outlined in [Andersen and Piterbarg (2010), p. 99]. For the noncentral $\chi^2$-distribution it is very expensive to use the quantile function as this involves numerical root search. We will elaborate on this later on when working with the Cox-Ingersoll-Ross (CIR) model, where the conditional distribution of the short rate is non-central $\chi^2$.

### 2.3.2 Discretisation Schemes

To get the simulation to work we need to discretise time. Suppose we have a SDE on the form

$$\mathrm{d}X(t) = \mu(t, X(t))\,\mathrm{d}t + \sigma(t, X(t))\,\mathrm{d}W(t) \tag{2.21}$$

The Euler scheme on an equidistant time grid would then be

$$\hat{X}(t+\Delta t) = \hat{X}(t) + \mu(\Delta t, \hat{X}(t))\Delta t + \sigma(\Delta t, \hat{X}(t))(W(t+\Delta t) - W(t)) \tag{2.22}$$

This scheme can however have stability issues. When $X(t)$ is Gaussian we can use the slightly different scheme

$$\hat{X}(t+\Delta t) = \hat{X}(t) + \mathbb{E}\left[X(t+\Delta t) - X(t)|\mathscr{F}_t\right] + \sqrt{\mathrm{Var}\left[X(t+\Delta t) - X(t)|\mathscr{F}_t\right]}Z \tag{2.23}$$

[4]Low-discrepancy sequences such as Sobol are quite efficient, but not of much use in our case, where the number steps per path are not fixed due to the possibility of default, which would 'waste' some random numbers generated by the Sobol sequence. Generally speaking the Monte Carlo simulation converges much faster using low-discrepancy sequences.

where $Z$ is standard normal. This is an exact simulation of the SDE.

An issue with the Euler scheme when the variable is not Gaussian is that there is a non-zero probability that the process will turn negative as the increments are Gaussian. This can be a problem when working with default probabilities, which obviously cannot be negative[5]. We will elaborate on discretisation schemes when discussing the implementation of concrete Monte-Carlo simulations. We can generate correlated Gaussian variables by using the Cholesky decomposition. Define a lower triangular matrix C such that

$$CC^{\top} = \Sigma \tag{2.24}$$

When $X$ is standard normal then $Z$ defined below will be normally distributed with mean $\mu$ and covariance matrix $\Sigma$

$$Z = \mu + C \cdot X \tag{2.25}$$

An example of how this would work in practice is shown below

**Example 2.12.** Given the covariance matrix we can calculate the Cholesky decomposition

$$\Sigma = \begin{pmatrix} 1 & 0 & 0.5 \\ 0 & 1 & -0.5 \\ 0.5 & -0.5 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0.5 & -0.5 & \frac{1}{\sqrt{2}} \end{pmatrix}$$

With $\mu = 0$ equation 2.25 can be used to generate correlated normal variables. This is illustrated below

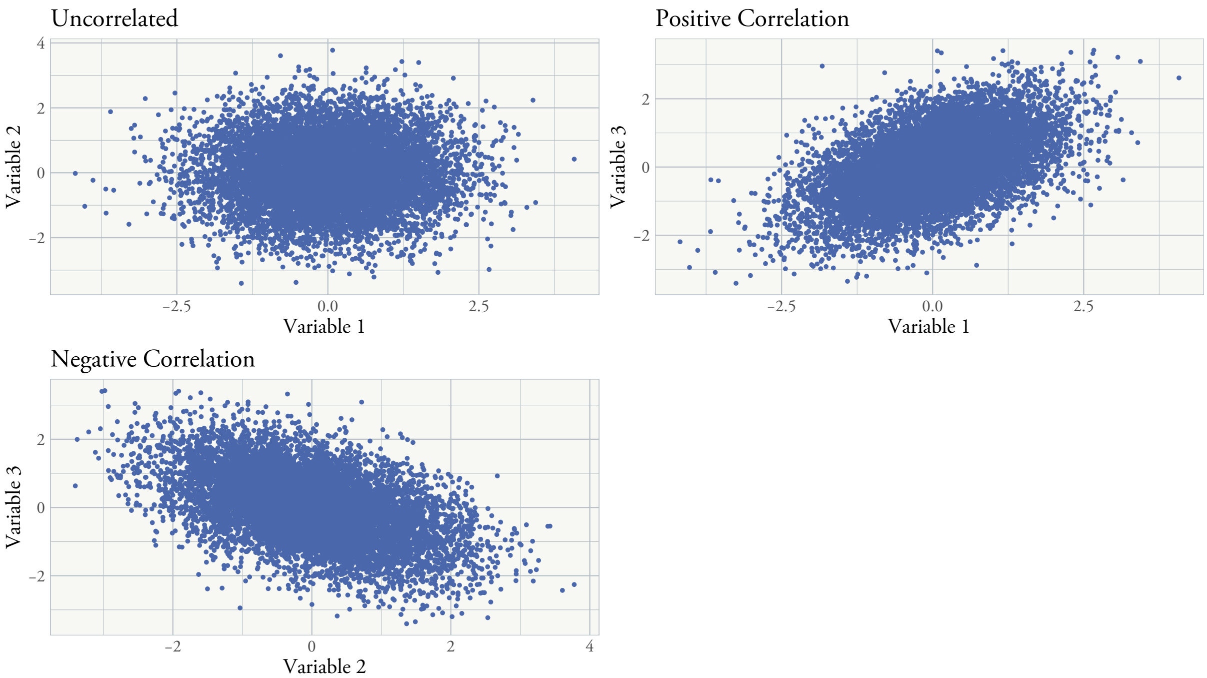



Figure 2.1: Generating Correlated Normals with Cholesky Decomposition

Generating correlated normals will be very important when introducing wrong and right way risk.

[5]Actually we model the hazard rate, but a negative hazard rate over some interval imply negative default probability over said interval.

# 3

# Interest Rate Modelling

In this chapter we will introduce relevant Fixed Income Derivatives and models used for pricing these. We will in general base our approach on [Andersen and Piterbarg (2010)] and will refer the reader to this book for further information on the topic.

In this chapter we will use the notation $\mathbb{E}_t[X(T)] := \mathbb{E}[X(T)|\mathscr{F}_t]$, where $\mathscr{F}_t$ is the natural filtration generated by the (possibly vector) Wiener process. As we are always using this filtration in this chapter, it should not cause any confusion. In later chapters, where we are working with multiple filtrations, we will of course state explicitly what filtration is used in the conditional expectation.

## 3.1 Bonds and Rates

The simplest claim one can think of is the *Zero Coupon Bond* (ZCB): $P(t,T)$. The owner is entitled to one Euro at maturity and no other payments are made to the bond owner. The maturity will typically be denoted $T$ and the time to maturity will then be $T-t$. This claim is also known as a discount bond as the bond can be seen as a discounting factor and it can be used as a 'building block' for pricing other fixed income instruments. Set $T_1 = T$ and $T_2 = T + \Delta t$ with $\Delta t > 0$. It can for example be used to price forward-starting ZCBs:

$$P(t;T_1,T_2) = \frac{P(t,T_2)}{P(t,T_1)} \tag{3.1}$$

This can be realised from the following trading strategy: buy 1 of the $T_2$-ZCB and sell short $\frac{P(t,T_2)}{P(t,T_1)}$ of the $T_1$-ZCB. Both these quantities are observable at time $t$. The cash flows can be observed in Table 3.1

| Time | Cash Flow |
|---|---|
| $t$ | $-1 \cdot P(t,T_2) + \frac{P(t,T_2)}{P(t,T_1)} P(t,T_1) = 0$ |
| $T_1$ | $-\frac{P(t,T_2)}{P(t,T_1)}$ |
| $T_2$ | $+1$ |

Table 3.1: Forward Starting ZCB

It is an arbitrage opportunity if the forward starting ZCB has a different price than the portfolio $\frac{P(t,T_2)}{P(t,T_1)}$ as they have the exact same payoff. It is completely equivalent to think of a continuously compounded discount rate

$$y(t;T_1,T_2) = -\frac{\ln P(t;T_1,T_2)}{\Delta t} \tag{3.2}$$

such that $y(t;T_{,1},T_2)$ solves $P(t;T_1,T_2) = e^{-y(t;T_1,T_2)\Delta t}$. Rates are always quoted on a yearly basis, so $\Delta t$ is the time in years from $T_1$ to $T_2$. $\Delta t$ is typically called the *tenor* of the bond. Rates are typically not continuously compounded, but compounded discretely. The simple forward rate is given by $F(t;T_1,T_2)$ in the equation

$$1 + \Delta t F(t;T_1,T_2) = \frac{1}{P(t;T_1,T_2)}$$

The left hand side generates interest on unit notional from time $T_1$ to time $T_2$ using the simple forward rate. The right hand side generates interest on unit notional from time $T_1$ to time $T_2$ using the forward starting ZCB. By the no-arbitrage principle the equality must hold. It is possible to get an expression for the simple forward rate

$$F(t;T_1,T_2) = \frac{1}{\Delta t}\left(\frac{1}{P(t;T_1,T_2)} - 1\right) \tag{3.3}$$

$F(T;T,T+\Delta t)$ will in this thesis be the *Euro Interbank Offered Rate* (Euribor) rate, i.e. the floating rate in a Euro swap. This rate is fixed in advance and paid in arrears. Historically this rate has been called $L$ standing for the London Interbank Offered Rate (Libor), but due to the Libor scandal this rate is being discontinued, which is also the reason for not choosing that notation. Not to be confused with the Futures rate though. The instantaneous forward rate is obtained in the limit:

$$f(t,T) := \lim_{\Delta t \downarrow 0} F(t;T_1,T_2) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t}\left(\frac{P(t,T) - P(t,T+\Delta t)}{P(t,T+\Delta t)}\right) = -\frac{\partial \ln P(t,T)}{\partial T} \tag{3.4}$$

We can use this rate to calculate forward starting ZCB prices as

$$P(t;T_1,T_2) = e^{-\int_{T_1}^{T_2} f(t,s)\,\mathrm{d}s} \tag{3.5}$$

Let $r(t) := f(t,t)$ be the short rate, i.e. the 'forward' rate from spanning $[t, t + \mathrm{d}t]$, which can loosely be thought of as an overnight rate.

## 3.2 Interest Rate Swaps

An interest rate swap is the exchange of a stream of fixed-rate payments for a stream of floating-rate payments. There are two types of swaps; one where you *receive* the fixed-rate payments and one where you *pay* the fixed-rate payments. They are called *receiver* and *payer* swaps, respectively. For swaps the tenor ($\Delta t$ from last section) is called year-fraction or *coverage*. There is quite a few day counting conventions for the coverage, which makes calculating the coverage non-trivial, but still somewhat simple. The dates on which the stream of payments fall are called a *schedule*. In the schedule we have to adjust for weekends and holidays as payments cannot fall on these days. There is a few date adjustment rules with the most common being *modified following*, which adjusts to the next business day unless this is in another month in which case it adjusts back to the previous business day. When the swap follows standard market conventions it is called *plain vanilla*[6].

If we for instance want to trade a spot-starting swap in Euro the start would be in two business days (spot $= t + 2$) with yearly fixed-rate payments, where coverage is calculated using the 30/360 convention and with floating-rate payments every 3 months, where coverage is calculated using the Act/360 convention. We would be adjusting for business days using the modified following convention. The 30/360 assumes 30 days in a month and 360 days in a year, while the Act/360 convention uses the actual amount of days and also assumes 360 days in a year.

An index $i$ will be used to keep track of the payment dates $T_i$. For payments related to the period between

[6]The term stems from the fact that vanilla is the most common ice cream flavour.

$T_{i-1}$ and $T_i$, where $T_{i-1} < T_i$, we denote the fixed rate $K$ and the floating rate $F(T_{i-1}; T_{i-1}, T_i)$ and the corresponding coverage $\delta^j(T_{i-1}, T_j)$, where $j \in \{K, L\}$. This gives us the fixed rate payment at time $T_i$

$$\delta^K(T_{i-1}, T_i)K$$

and the floating rate payment at time $T_i$

$$\delta^F(T_{i-1}, T_i)F(T_{i-1}; T_{i-1}, T_i)$$

To ease of notation define the shorthand notation $\delta_i^j := \delta^j(T_{i-1}, T_i)$. Let $N_K$ denote the number of fixed rate payments and let $N_L$ be the number of floating rate payments. The swap starts in $T_0$. The time $T_i$ fixed rate payment can be priced as

$$\beta(t)\mathbb{E}_t^{\mathbf{Q}}\left[\delta_i^K K \frac{1}{\beta(T_i)}\right] = \delta_i^K K \mathbb{E}_t^{\mathbf{Q}}\left[\frac{\beta(t)}{\beta(T_i)}\right] = \delta_i^K K P(t, T_i)$$

as $\delta_i^K$ and $K$ are constants. For the time $T_i$ floating rate payment we will use the forward measure, where $P(t, T_i)$ is numeraire, for pricing

$$P(t, T_i)\mathbb{E}_t^{\mathbf{T_i}}\left[\delta_i^F F(T_{i-1}; T_{i-1}, T_i)\frac{1}{P(T_i, T_i)}\right] = P(t, T_i)\delta_i^F F(t; T_{i-1}, T_i)$$

as $P(T_i, T_i) = 1$ and due to the forward rate being a martingale as $\frac{1}{P(t;T_{i-1},T_i)} = \frac{P(t,T_{i-1})}{P(t,T_i)}$ is a martingale under $\mathbf{T_i}$. Let the fixed rate payment schedule be given by

$$i^K(t) := \{i : t < T_i \leq T_{N_K}\} \tag{3.6}$$

and the floating rate payment schedule be given by

$$i^F(t) := \{i : t < T_i \leq T_{N_L}\} \tag{3.7}$$

We can then write the fair price of a payer swap as

$$V(t) = \sum_{i^F(t)} \delta_i^F F(t; T_{i-1}, T_i)P(t, T_i) - \sum_{i^K(t)} \delta_i^K K P(t, T_i) \tag{3.8}$$

Note that $-V(t)$ is the value of an equivalent receiver swap. Unit notional has been assumed. Usually we are only interested in the unrealised part of the swap, hence the $t > T_i$ in the schedules. The pricing of a realised amount is trivial and does not constitute a market or credit risk. Note that the $T_i$s will differ in the two sums, if the payment schedules are not equal for the two legs of the swap, which they usually are not. The price of a subset of the swap will be noted by

$$V(t, T) = V(t)\mathbb{1}_{\{T_i<T\}} \tag{3.9}$$

where $T$ is the last time of interest. It should be very clear now that we only need prices on ZCBs today to price a swap without credit risk. Given market prices for ZCB's the pricing of a swap will be model independent. If there does not exist ZCB prices for all the relevant dates, then a proper interpolation method will be needed[7], but still the valuation of a swap does not depend on many assumptions.

Usually a swap is entered at-market meaning that the net present value of the swap is zero initially. A relevant concept in this regard is the *par swap rate* that makes the swap costless. This is also called the fair rate for the swap. This can easily be found

$$0 = \sum_{i^K(t)} \delta_i^K S(t) P(t, T_i) - \sum_{i^F(t)} \delta_i^F F(t; T_{i-1}, T_i) P(t, T_i)$$

$$\Longleftrightarrow$$

$$S(t) := \frac{\sum_{i^F(t)} \delta_i^F F(t; T_{i-1}, T_i) P(t, T_i)}{\sum_{i^K(t)} \delta_i^K P(t, T_i)} \tag{3.10}$$

Note that the par swap rate is the same for payer and receiver swaps. Let the *annuity* be

$$A(t) := \sum_{i^K(t)} \delta_i^K P(t, T_i) \tag{3.11}$$

The annuity is the present value of one basis point change in interest rates. The notation for the price of a payer swap can then be simplified to

$$V(t) = A(t)[S(t) - K] \tag{3.12}$$

The representation above gives us the idea that we can value the swap under the so-called *annuity measure* that we denote **A**. This is possible as the annuity, as a weighted sum of discounting factors, is always positive. The pricing formula for the swap starting at $T_0$ is given by

$$V(t) = A(t)\mathbb{E}_t^{\mathbf{A}}\left[A(T_0)\,(S(T_0) - K)\,\frac{1}{A(T_0)}\right] = A(t)[S(t) - K] \tag{3.13}$$

where the last equality holds as the forward swap rate is a martingale under the annuity measure: $S(t)$ has $A(t)$ in the denominator.

Going forward we will denote the value of a payer swap $V_P(t)$ and the value of a receiver swap $V_R(t)$ to avoid confusion. We will denote the value of a payer swaption as $V_{P+}(t)$ and the value of a receiver swaption $V_{R+}(t)$.

### 3.2.1 Dual Curve Setup

Until now we have implicitly assumed that the same interest rate curve can be used to construct both discounting rates and forecasting future rates. Before the financial crisis of 2007-2008 this was a good approximation as banks were in general able to fund itself at USD Libor rates, so the Libor curve was also used for discounting. There was a spread between the overnight curve and Libor and also basis in the Cross Currency Swaps, but the spreads were in general too small to be arbirtrageable. After the credit crunch, liquidity dried up and spreads widened dramatically, making it obvious that Libor could not be used for discounting anymore. The appropriate discounting rate is the *overnight rate* as this is the rate that the bank can use for funding in the very short term (literally overnight). This is a risk-free rate in contrast to Libor, which includes the credit risk of the member banks.

[7] In this case, *proper* means an interpolation method that does not introduce any arbitrage opportunities into the pricing model. This is discussed in chapter 6 of [Andersen and Piterbarg (2010)].

Additionally the Libor scandal have brought issues to light. There has been widespread manipulation with Libor rates, which is the reason for the ongoing IBOR transition, where alternative reference rates are being developed. These new rates have different methodology and this thesis will focus on the Euro rates, but information on other currencies can be found at this link. The unsecured risk-free rate *Euro Short Term Rate* (€STR) will replace the current *Euro Overnight Index Average* (Eonia) and Eonia is now calculated as €STR plus 8.5 basis points, but will be discontinued on 3rd January 2022. €STR is a backwards-looking rate and it is based on actual transactions (overnight borrowing between banks as the name suggests) [European Central Bank (2020)]. There is at the moment no plan to end *Euro Interbank Offered Rate* (Euribor) and today it is also based on actual transactions.

To calibrate the discounting curve a liquid instrument that depend on the overnight rate is needed. This is the *Overnight Index Swap* (OIS) described below

**Definition 3.1 (Overnight Index Swap).** An OIS is a fixed-for-floating interest rate swap, where the floating rate is the compounded overnight rate.

Let the overnight rate compound over $n$ days between $T_0$ and $T_n$ such that $\{T_i\}_{i\in\{1,2,\dots,n\}}$ contains all business days in the period. Let $F(T_i;T_{i-1},T_i)$ be the overnight rate between $T_{i-1}$ and $T_i$. Notice that this is a time $\mathscr{F}_{T_i}$-measurable variable, i.e. only observable at the end of the period. $\delta$ is the coverage and it is assumed equal for both the fixed rate and the floating rate.

$$\overline{F}(T_n;T_0,T_n) = \frac{1}{\delta(T_0,T_n)}\left(\prod_{i=1}^{n}(1+\delta_{i-1}F(T_i;T_{i-1},T_i)) - 1\right) \tag{3.14}$$

Assuming continuous compounding the expression will be the same as the simple forward rate $\overline{F}(T_n;T_0,T_n) = F(T_n;T_0,T_n)$, although this is not entirely correct. However, this approximation would make it possible to use the same approach as for the swap on a simple forward rate. The payoff of the payer OIS is

$$\delta(T_0,T_n)\left(\overline{F}(T_n;T_0,T_n) - K\right) \tag{3.15}$$

This is equivalent to a *swaplet* with the same coverage for the floating rate and the fixed rate.

Right now the spread between 3m Euribor and €STR is around 3 bp, so this is deemed insignificant. Since its inception in October 2019 €STR has been rather stable with a high of -0.511% on 20th November 2019 and a low of -0.583% on 31st December 2020. The 3m Euibor on the other hand has a high of -0.151% on 23rd April 2020 and a low of -0.556% on 7th January 2021 within the same period[8]. It is clear that Euribor is far more volatile than €STR, which also suggests that €STR works as a risk-free rate as it is not affected by short term changes in liquidity as was the case for Euribor in Spring 2020.

The reason for not using a dual curve setup in this thesis is that it would make calibration and simulation far more complicated. With the current spread between 3m Euribor and €STR it would also probably not have a large effect on CVA and DVA.

If the trade is collateralised the appropriate funding rate is OIS as collateral earns interest based on the overnight rate, but if the trade is uncollateralised a Funding Value Adjustment might be needed to account for a different funding rate for the bank.

### 3.2.2 Risk Management

The risk of changes in interest rates is the most relevant for the risk management of linear interest rate derivatives. The first step in determining the risk of a swap portfolio is the construction of the yield curve.

[8]Data on €STR is available in ECB's Statistical Data Warehouse here and data on Euribor is available here.

The yield curve is constructed by bootstrapping to a set of market instruments. In this thesis swap rates are the only instrument used for determining zero rates, but usually deposit rates and *Forward Rate Agreements* (FRAs) are used for the short end of the curve, i.e. for maturities less than one year. It is important that the model rates are consistent with the rates observed in the market to avoid the prices implied by the model being arbitrageable in terms of the true prices. To exemplify, if a bank used a model that was not calibrated to market prices, it might offer a price too low or bid a price that is too high. This would be bad for business.

Define the $n$-dimensional model parameter vector as $\Psi$, which is the zero rates. $\Psi$ will occasionally be ZCBs as the bond prices are equivalent with the rates. In the bootstrapping we have to minimise the squared distance between model and market prices by changing the model parameters. Mathematically this can be written as

$$\min_{\Psi} ||\text{Model}(\Psi) - \text{Market}||^2 \tag{3.16}$$

It will typically be a good idea to choose the same number of model parameters as there is market prices as this ensures that an exact solution is found. This can be seen from the first order condition

$$(\text{Model}(\Psi) - \text{Market})^\top \frac{\partial \text{Model}(\Psi)}{\partial \Psi} = 0 \tag{3.17}$$

where $\frac{\partial \text{Model}(\Psi)}{\partial \Psi}$ is the Jacobian matrix. As the Jacobian is non-singular when it is square, the problem can be simplified to

$$\text{Model}(\Psi) - \text{Market} = 0 \tag{3.18}$$

This a system of $n$ equations in $n$ unknowns, so it has exactly one solution. The $n$ model parameters has now been determined. As mentioned earlier a proper interpolation method is needed to determine the rate on a date that is not one of the $n$ dates. The interpolation method of choice will typically be *Cubic Splines* as this ensures that both zero rates and forward rates are smooth. If there are issues with the cubic interpolation then a linear interpolation is chosen. The reader is referred to [Andersen and Piterbarg (2010), pp. 240–243] for more information on this. Other interpolation methods are explained as well. Extrapolation of rates is always flat.

A risk measure widely used in Fixed Income is the *Dollar Value of 1 basis point* (DV01). It is the change in portfolio value given a parallel upwards shift in the yield curve of 1 basis point. For a portfolio with value process $\Pi$ this can be written as

$$DV01(\Pi) := \Pi(\Psi + 1bp) - \Pi(\Psi) \approx \frac{1}{10\,000} \frac{\partial \Pi(\Psi)}{\partial \Psi} \tag{3.19}$$

where $\frac{\partial \Pi(\Psi)}{\partial \Psi}$ is the *model delta vector*. The model delta vector shows which zero rates the portfolio is sensitive to and by how much. This approach is called the *bump and re-value* approach, which is essentially a low-tech description of a finite difference method. For hedging purposes however, the model delta vector is not relevant, but rather the *market delta vector*, which is the sensitivity of the portfolio towards changes in market prices – typically swap rates.

**Proposition 3.2 (Market Delta Vector).** *The market delta vector can be determined from the model delta vector and the Jacobian as*

$$\frac{\partial \Pi(\Psi)}{\partial Market} = \left( \frac{\partial Model(\Psi)}{\partial \Psi}^\top \right)^{-1} \frac{\partial \Pi(\Psi)}{\partial \Psi}^\top \tag{3.20}$$

*Proof.* Using the chain rule of multivariate calculus we can write

$$\frac{\partial \Pi(\Psi)}{\partial \Psi} = \frac{\partial \Pi(\Psi)}{\partial \mathrm{Model}}^{\top} \frac{\partial \mathrm{Model}(\Psi)}{\partial \Psi} \tag{3.21}$$

Transposing both sides of the equality yields

$$\frac{\partial \Pi(\Psi)}{\partial \Psi}^{\top} = \frac{\partial \mathrm{Model}(\Psi)}{\partial \Psi}^{\top} \frac{\partial \Pi(\Psi)}{\partial \mathrm{Model}} \tag{3.22}$$

where we can easily isolate

$$\left( \frac{\partial \mathrm{Model}(\Psi)}{\partial \Psi}^{\top} \right)^{-1} \frac{\partial \Pi(\Psi)}{\partial \Psi}^{\top} = \frac{\partial \Pi(\Psi)}{\partial \mathrm{Model}} \tag{3.23}$$

As the model has been calibrated to market prices such that $\mathrm{Model}(\Psi) = \mathrm{Market}$, we have now arrived at the market delta vector. ■

When the market delta vector is calculated, hedging of the portfolio can be done by trading the market instruments, where the delta value is the largest.

## 3.3 Heath-Jarrow-Morton Framework

The *Heath-Jarrow-Morton* (HJM) framework is a very general approach to interest rate modelling. It evolves around describing the dynamics of a continuum of ZCB's with finite maturity. It takes the current market prices for ZCB's as input and the dynamics of the yield curve can then be set arbitrarily up to some regularity conditions. This is the modern approach to interest rate modelling. The bond discounted by the bank account will be denoted by

$$P_\beta(t,T) = \frac{P(t,T)}{\beta(t)} \tag{3.24}$$

As this is a tradable asset, this will be a martingale under $\mathbf{Q}$

$$\mathrm{d}P_\beta(t,T) = -P_\beta(t,T)\sigma_P(t,T)^{\top}\,\mathrm{d}W^{\mathbf{Q}}(t) \tag{3.25}$$

where the yield curve is driven by a $d$-dimensional Wiener process. The reason for putting a minus in front of the volatility is that the forward rate volatility and bond volatility are inversely related due to the inverse relationship between bond prices and rates. The condition $\sigma_P(T,T) = 0$ makes the price 'pull to par', i.e. be 1 at maturity. Obviously the bond itself will have local return rate $r(t)\,\mathrm{d}t$ under $\mathbf{Q}$

$$\frac{\mathrm{d}P(t,T)}{P(t,T)} = r(t)\,\mathrm{d}t - \sigma_P(t,T)^{\top}\,\mathrm{d}W^{\mathbf{Q}}(t) \tag{3.26}$$

Using Itô's lemma we can find the dynamics of the forward bond, where $T_1$ and $T_2$ are defined in the same manner as earlier

$$
\begin{aligned}
\mathrm{d}\left(\frac{P(t,T_2)}{P(t,T_1)}\right) &= \frac{1}{P(t,T_1)}\,\mathrm{d}P(t,T_2) - \frac{P(t,T_2)}{P(t,T_1)^2}\,\mathrm{d}P(t,T_1) \\
&\quad - \frac{1}{P(t,T_1)^2}\,\mathrm{d}P(t,T_2)\,\mathrm{d}P(t,T_1) + \frac{P(t,T_2)}{P(t,T_1)^3}(\mathrm{d}P(t,T_1))^2
\end{aligned}
$$

which can be simplified to

$$
\begin{aligned}
\mathrm{d}\left(\frac{P(t,T_2)}{P(t,T_1)}\right) &= \frac{P(t,T_2)}{P(t,T_1)}\left(-\sigma_P(t,T_2)^\top\,\mathrm{d}W^{\mathbf{Q}}(t) + \sigma_P(t,T_1)^\top\,\mathrm{d}W^{\mathbf{Q}}(t)\right) \\
&\quad + \frac{P(t,T_2)}{P(t,T_1)}\left(-\sigma_P(t,T_2)^\top\sigma_P(t,T_1)\,\mathrm{d}t + \sigma_P(t,T_1)^\top\sigma_P(t,T_1)\,\mathrm{d}t\right)
\end{aligned}
$$

We can thus write

$$
\frac{\mathrm{d}P(t;T_1,T_2)}{P(t;T_1,T_2)} = -\left[\sigma_P(t,T_2) - \sigma_P(t,T_1)\right]^\top \sigma_P(t,T_1)\,\mathrm{d}t - \left[\sigma_P(t,T_2) - \sigma_P(t,T_1)\right]^\top \mathrm{d}W^{\mathbf{Q}}(t) \tag{3.27}
$$

Under the $T_1$-forward measure $\mathbf{T_1}$, the forward ZCB is a martingale

$$
\frac{\mathrm{d}P(t;T_1,T_2)}{P(t;T_1,T_2)} = -\left[\sigma_P(t,T_2) - \sigma_P(t,T_1)\right]^\top \mathrm{d}W^{\mathbf{T_1}}(t) \tag{3.28}
$$

Using Itô's lemma we can also find the dynamics of the logarithm to the bond

$$
\mathrm{d}\ln P(t,T) = \left(r(t) + \frac{1}{2}\sigma_P(t,T)^\top\sigma_P(t,T)\right)\mathrm{d}t - \sigma_P(t,T)^\top\,\mathrm{d}W^{\mathbf{Q}}(t) \tag{3.29}
$$

The drift is not really of importance – the important part is the diffusion as we need this to determine the dynamics of the instantaneous forward rate. As $f(t,T) = -\frac{\partial \ln P(t,T)}{\partial T}$ the relationship $\sigma_f(t,T) = \frac{\partial}{\partial T}\sigma_P(t,T)$ is established.

Traditionally HJM analysis begin with defining the dynamics of the instantaneous forward rate. In all generality we can write the dynamics as

$$
\mathrm{d}f(t,T) \coloneqq \mu_f(t,T)\,\mathrm{d}t + \sigma_f(t,T)^\top\,\mathrm{d}W^{\mathbf{Q}}(t) \tag{3.30}
$$

To ensure no arbitrage in the model we must restrict the drift term

**Proposition 3.3 (HJM Drift Condition).** *To ensure no arbitrage in the HJM framework we must have*

$$
\mu_f(t,T) = \sigma_f(t,T)^\top \int_t^T \sigma_f(t,s)\,\mathrm{d}s \tag{3.31}
$$

*or equivalently*

$$
\mu_f(t,T) = \sigma_f(t,T)^\top\sigma_P(t,T) \tag{3.32}
$$

*Proof.* Under the forward measure: $f(t,T)$ is a martingale. This can be seen from equation 3.4.

$$\begin{aligned} \mathrm{d}f(t,T) &= \sigma_f(t,T)^\top \,\mathrm{d}W^{\mathbf{T}}(t) \\ &= \sigma_f(t,T)^\top \left(\mathrm{d}W^{\mathbf{Q}}(t) + \sigma_P(t,T)\,\mathrm{d}t\right) \\ &= \sigma_f(t,T)^\top \sigma_P(t,T)\,\mathrm{d}t + \sigma_f(t,T)^\top \,\mathrm{d}W^{\mathbf{Q}}(t) \\ &= \sigma_f(t,T)^\top \int_t^T \sigma_f(t,s)\,\mathrm{d}s\,\mathrm{d}t + \sigma_f(t,T)^\top \,\mathrm{d}W^{\mathbf{Q}}(t) \end{aligned} \tag{3.33}$$

Which shows that the drift must be $\sigma_f(t,T)^\top \int_t^T \sigma_f(t,s)\,\mathrm{d}s$ to ensure that the instantaneous forward rate is a martingale under $\mathbf{T}$. ■

This is considered the main result of the HJM framework. It shows that we only need to specify the volatility to have a model with no arbitrage. The $f(0,T)$ is exogenous and the model will thus match all observed ZCB prices, no matter how $\sigma_f(t,T)$ is set (up to some regularity constraints). This is a very nice result, but far too general to work in practice. In its most general form the continuum of forward rates are an infinite-dimensional diffusion process. To get it work in practice we can make specific assumptions such that it become a Markovian Gaussian model.

We will later be interested in models for the short rate. In the HJM framework the short rate is given by

$$r(t) = f(t,t) = f(0,t) + \int_0^t \sigma_f(u,t)^\top \int_u^t \sigma_f(u,s)\,\mathrm{d}s\,\mathrm{d}u + \int_0^t \sigma_f(u,t)^\top \,\mathrm{d}W^{\mathbf{Q}}(u) \tag{3.34}$$

This is in general not Markovian as the diffusion is path-dependent. Set

$$D(t) = \int_0^t \sigma_f(u,t)^\top \,\mathrm{d}W^{\mathbf{Q}}(u)$$

so

$$\begin{aligned} D(T) &= \int_0^T \sigma_f(u,T)^\top \,\mathrm{d}W^{\mathbf{Q}}(u) \\ &= D(t) - \int_0^t \sigma_f(u,t)^\top \,\mathrm{d}W^{\mathbf{Q}}(u) \\ &\quad + \int_0^t \sigma_f(u,T)^\top \,\mathrm{d}W^{\mathbf{Q}}(u) + \int_t^T \sigma_f(u,T)^\top \,\mathrm{d}W^{\mathbf{Q}}(u) \end{aligned}$$

It is clear that the value of $D(T)$ depends on the path from 0 to $t$ and not only $D(t)$ and it is thus not Markov in general. Assuming that $\sigma_f(t,T)$ is a bounded and deterministic function of $t$ and $T$ then the short rate is Gaussian with

$$\mathbb{E}_t^{\mathbf{Q}}[r(T)] = f(t,T) + \int_t^T \sigma_f(u,T)^\top \int_u^T \sigma_f(u,s)\,\mathrm{d}s\,\mathrm{d}u \tag{3.35}$$

$$\mathrm{Var}_t^{\mathbf{Q}}[r(T)] = \int_t^T \sigma_f(u,T)^\top \sigma_f(u,T) \tag{3.36}$$

If volatility is separable then it is also Markovian. Consider the special choice

$$\sigma_f(t,T) = g(t)h(T)$$

Then

$$\sigma_P(t,T) = \int_t^T \sigma_f(t,s)\,\mathrm{d}s = g(t)\int_t^T h(s)\,\mathrm{d}s$$

and consequently

$$r(t) = f(0,t) + h(t)\int_0^t g(u)^\top g(u)\left(\int_u^t h(s)\,\mathrm{d}s\right)\mathrm{d}u + h(t)\int_0^t g(u)^\top\,\mathrm{d}W^{\mathbf{Q}}(u) \tag{3.37}$$

We now have

$$D(t) = h(t)\int_0^t g(u)^\top\,\mathrm{d}W^{\mathbf{Q}}(u)$$

which makes it possible to write

$$\begin{aligned}
D(T) &= h(T)\int_0^T g(u)^\top\,\mathrm{d}W^{\mathbf{Q}}(u)\\
&= \frac{h(T)}{h(t)}\left(h(t)\int_0^t g(u)^\top\,\mathrm{d}W^{\mathbf{Q}}(u)\right) + h(T)\int_t^T g(u)^\top\,\mathrm{d}W^{\mathbf{Q}}(u)\\
&= \frac{h(T)}{h(t)}D(t) + h(T)\int_t^T g(u)^\top\,\mathrm{d}W^{\mathbf{Q}}(u)
\end{aligned}$$

This does not depend on the path between 0 and $t$, but only $D(t)$, so it is Markovian.

**Proposition 3.4 (HJM with Gaussian and Markov Short Rate).** *The short rate specification*

$$\mathrm{d}r(t) = (a(t) - \kappa(t)r(t))\,\mathrm{d}t + \sigma_r(t)^\top\,\mathrm{d}W^{\mathbf{Q}}(t) \tag{3.38}$$

*is consistent with the HJM framework, when $\kappa$ and $\sigma_r$ are deterministic and*

$$a(t) = \frac{\partial f(0,t)}{\partial t} + \kappa(t)f(0,t) + \int_0^t e^{-2\int_u^t \kappa(s)\,\mathrm{d}s}\sigma_r(u)^\top\sigma_r(u)\,\mathrm{d}u \tag{3.39}$$

*Proof.* Set

$$h(T) = e^{-\int_0^T \kappa(s)\,\mathrm{d}s},\quad g(t) = e^{\int_0^t \kappa(s)\,\mathrm{d}s}\sigma_r(t)$$

This means

$$\frac{h'(t)}{h(t)} = -\kappa(t)$$

and

$$\sigma_f(t,T) = e^{\int_0^t \kappa(s)\,\mathrm{d}s}e^{-\int_0^T \kappa(s)\,\mathrm{d}s}\sigma_r(t) = e^{-\int_t^T \kappa(s)\,\mathrm{d}s}\sigma_r(t)$$

Equation (4.46) in [Andersen and Piterbarg (2010), p. 188] states

$$\begin{aligned}
\mathrm{d}r(t) &= \left(\frac{\partial f(0,t)}{\partial t} - \frac{h'(t)}{h(t)} f(0,t) + h(t)^2 \int_0^t g(u)^\top g(u)\,\mathrm{d}u + \frac{h'(t)}{h(t)} r(t)\right)\mathrm{d}t \\
&\quad + h(t)g(t)^\top\,\mathrm{d}W^{\mathbf{Q}}(t)
\end{aligned}$$

This expression is found by differentiating equation 3.37 with respect to $t$. In this case it becomes

$$\begin{aligned}
\mathrm{d}r(t) &= \left(\frac{\partial f(0,t)}{\partial t} + \kappa(t) f(0,t) + e^{-2\int_0^t \kappa(s)\,\mathrm{d}s} \int_0^t e^{2\int_0^u \kappa(s)\,\mathrm{d}s} \sigma_r(u)^\top \sigma_r(u)\,\mathrm{d}u - \kappa(t) r(t)\right)\mathrm{d}t \\
&\quad + e^{-\int_0^t \kappa(s)\,\mathrm{d}s} e^{\int_0^t \kappa(s)\,\mathrm{d}s} \sigma_r\,\mathrm{d}W^{\mathbf{Q}}(t) \\
&= \left(\frac{\partial f(0,t)}{\partial t} + \kappa(t) f(0,t) + \int_0^t e^{-2\int_u^t \kappa(s)\,\mathrm{d}s} \sigma_r(u)^\top \sigma_r(u)\,\mathrm{d}u - \kappa(t) r(t)\right)\mathrm{d}t \\
&\quad + \sigma_r\,\mathrm{d}W^{\mathbf{Q}}(t) \\
&= (a(t) - \kappa(t) r(t))\,\mathrm{d}t + \sigma_r\,\mathrm{d}W^{\mathbf{Q}}(t)
\end{aligned}$$

■

In the Gaussian HJM model we have

$$\sigma_f(t,T) = e^{-\int_0^T \kappa(s)\,\mathrm{d}s} e^{\int_0^t \kappa(s)\,\mathrm{d}s} \sigma_r(t) = e^{-\int_t^T \kappa(s)\,\mathrm{d}s} \sigma_r(t) \tag{3.40}$$

and

$$\sigma_P(t,T) = e^{\int_0^t \kappa(s)\,\mathrm{d}s} \sigma_r(t) \int_t^T e^{-\int_0^u \kappa(s)\,\mathrm{d}s}\,\mathrm{d}u = \sigma_r(t) \int_t^T e^{-\int_t^u \kappa(s)\,\mathrm{d}s}\,\mathrm{d}u \tag{3.41}$$

**Proposition 3.5 (Option on ZCB).** *Assume that the short-rate is given by equation 3.34. Consider a European call option on a Zero Coupon Bond with payoff*

$$(P(T_E, T_M) - K)^+ \tag{3.42}$$

*where $T_E$ is the expiry of the option and $T_M$ is the maturity of the bond with $T_E < T_M$. The value of this option is given by*

$$C(t) = P(t, T_M)\Phi(d_+) - P(t, T_E) K \Phi(d_-) \tag{3.43}$$

*where*

$$d_\pm = \frac{\ln\left(\frac{P(t,T_M)}{P(t,T_E)K}\right) \pm \frac{v}{2}}{\sqrt{v}} \tag{3.44}$$

*and*

$$v = \int_t^T |\sigma_P(u, T_M) - \sigma_P(u, T_E)|^2\,\mathrm{d}u \tag{3.45}$$

*where*

$$|x|^2 := \mathrm{tr}(x \cdot x^\top) \tag{3.46}$$

*Proof.* Under the $T_E$ forward measure the forward ZCB $P(T_E; T_E, T_M)$ is a martingale. This is shown in equation 3.28. We can thus represent the price as

$$\begin{aligned} C(t) &= P(t, T_E)\mathbb{E}_t^{\mathbf{T_E}}\left[(P(T_E, T_M) - K)^+\right] \\ &= P(t, T_E)\mathbb{E}_t^{\mathbf{T_E}}\left[(P(T_E; T_E, T_M) - K)^+\right] \end{aligned}$$

The forward ZCB is log-normal with variance[9]

$$v = \int_t^{T_E} |\sigma_P(u, T_M) - \sigma_P(u, T_E)|^2 \, \mathrm{d}u$$

so we can use Black's formula from Proposition 3.6, where we substitute $S(t)$ for $P(t; T_E, T_M) = \frac{P(t,T_M)}{P(t,T_E)}$ and $\sigma\sqrt{T-t}$ for $v$, which immediately leads us to the formula

$$C(t) = P(t, T_E)\left(\frac{P(t, T_M)}{P(t, T_E)}\Phi(d_+) - K\Phi(d_-)\right) = P(t, T_M)\Phi(d_+) - P(t, T_E)K\Phi(d_-) \qquad \blacksquare$$

The value of a put option on a ZCB can be found by the put-call parity, which can be verified by the payoffs

$$(K - P(T_E, T_M))^+ = (P(T_E, T_M) - K)^+ - (P(T_E, T_M) - K)$$

Being long a put is the same as being long a call and being short an agreement to buy the forward ZCB for $K$.

The option on a ZCB plays the same role for non-linear interest rate derivatives as the ZCB plays for linear interest rate derivative. It can be seen as a 'building block'. We will elaborate on this in the next section.

## 3.4 Swaptions

A *European swaption* is the option to enter a swap starting at a future point in time. This means that you as a buyer of a swaption have the right, but not the obligation to enter the underlying swap. This protects you against downside risk, but will cost you a risk-premium. If the swap starts at time $T_0$ as was the case earlier, we can express the value of the payer swaption as

$$V_{P+}(T_0) = (V(T_0))^+ = (A(T_0)[S(T_0) - K])^+ = A(T_0)\,(S(T_0) - K)^+$$

where the last equality follows from the fact that $A(T_0) > 0$. Obviously the value at time $t < T_0$ is

$$V_{P+}(t) = \beta(t)\mathbb{E}_t^{\mathbf{Q}}\left[\frac{1}{\beta(T_0)}A(T_0)\,(S(T_0) - K)^+\right]$$

[9]The variance here refers to the variance of $X$, when $X$ is normal and $Y = e^X$.

or equivalently under the annuity measure

$$V_{P+}(t) = A(t)\mathbb{E}_t^{\mathbf{A}}\left[(S(T_0) - K)^+\right]$$

where the swap rate is a martingale.

### 3.4.1 Vanilla Models

The first attempt to price Swaptions will be in the Vanilla model framework, where assumptions about the swap rate dynamics are applied directly.

**Proposition 3.6 (Black formula).** *Assume that the par swap rate follows*

$$\mathrm{d}S(t) = \sigma S(t)\,\mathrm{d}W(t)$$

*where $W(t)$ is an* **A**-*Wiener process. The swaption expires at time $T$ (which must be before the start of the swap, which will be denoted $T_0$) and the underlying swap has fixed rate $K$. The time $t$ price of a payer swaption then becomes*

$$V_{P+}(t) = A(t)\left[S(t)\Phi(d_+) - K\Phi(d_-)\right]$$
$$d_\pm = \frac{\ln\frac{S(t)}{K} \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}$$

Proof is in the appendix A.2 as this should be straight forward.

The price of a receiver swaption follows from the put-call parity for swaptions

$$V_{P+}(t) - V_{R+}(t) = V_P(t)$$

which can be easily verified by their payoffs

$$A(T)(S(T) - K)^+ - A(T)(K - S(T))^+ = A(T)(S(T) - K)$$

So a receiver swaption can be replicated by a long position in a payer swaption as well as a short position in the underlying swap.

One obvious problem with the Black formula is that it does not allow for negative rates. As as it has been seen in recent years the interest rate can be negative, so the swap rates can also be negative. Usually swaption prices are quoted as an *implied Black volatility* as it is then easier to compare swaptions across strikes (moneyness) and expiries. Every parameter apart from $\sigma$ is observable in the market, so we can find the implied volatility by inverting the Black formula

$$\text{Black}^{-1}(\text{Observed Price}) = \text{Implied Black Volatility} \tag{3.47}$$

It should be noted that we cannot obtain an explicit expression for the inverse of the Black formula, so we have to use a numerical method to solve the equation. One alternative to the Black implied volatility, when rates are low or negative, is the *Gaussian implied volatility*, where we assume that the par swap rate's conditional distribution is normal instead of log-normal. In this case we can use the Bachelier formula for a call option in a similar way

**Proposition 3.7 (Bachelier formula).** *Assume that the par swap rate follows the SDE*

$$\mathrm{d}S(t) = \sigma\,\mathrm{d}W(t)$$

*where $W(t)$ is an* **A**-*Wiener process. The swaption expires at time $T$ (which must be before the start of the swap) and the underlying swap has fixed rate $K$. The time $t$ price of a payer swaption then becomes*

$$V_{P+}(t) = A(t)\left[(S(t) - K)\,\Phi(d) + \sigma\sqrt{T-t}\phi(d)\right]$$
$$d = \frac{S(T) - K}{\sigma\sqrt{T-t}}$$

The proof of this proposition is also in Appendix A.2.

We will not find the explicit price of a receiver swaption under this model as it again follows from the put-call parity of swaptions.

It should be clear that we can also find the Bachelier implied volatility by inverting the formula. In this model the $\sigma$ is the absolute volatility instead of a percentage of the swap rate. The Gaussian implied volatility is typically quoted in *basis points* (bp), which is 1 percent of 1 percent: $\frac{1}{10000}$.

More complicated vanilla models such as ZABR exist, which is an extension to SABR to allow for negative rates. These models are widely used in practice for risk management purposes, but typically with one model per maturity. This makes the model unsuitable for CVA purposes as joint evolution of a vector of ZABR models would be very expensive, computationally speaking.

### 3.4.2 Greeks

For non-linear interest rate derivatives it is not enough to calculate the delta vector. In terms of model sensitivities the typical approach is to determine these using the derivatives in the Black or Bachelier formula with respect to the sensitivity of interest. Let the pricing formula used be denoted by $F$. Then the partial derivatives can be used to measure the risk

$$\text{Theta: } \frac{\partial F}{\partial t} \quad \text{Delta: } \frac{\partial F}{\partial S} \quad \text{Gamma: } \frac{\partial^2 F}{\partial S^2} \quad \text{Vega: } \frac{\partial F}{\partial \sigma} \quad \text{Vanna: } \frac{\partial^2 F}{\partial \sigma \partial S} \quad \text{Volga: } \frac{\partial^2 F}{\partial \sigma^2}$$

The Greeks are thus found by calculating the partial derivatives of the model in question. In practice the Greeks for a portfolio can be found using finite difference methods.

A swaption straddle is a strategy that involves buying a payer and a receiver swaption at the same strike, expiry and maturity. At expiry ($T_0$) the payoff is given by

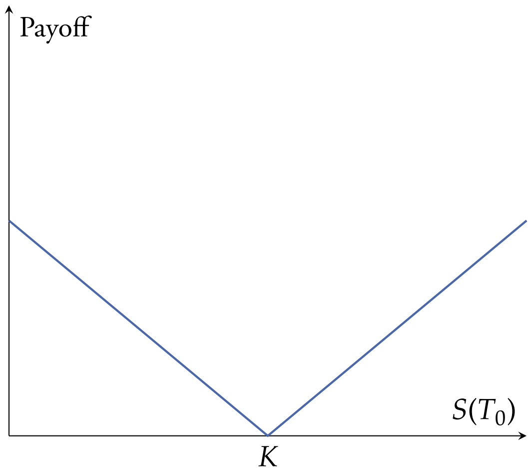



Figure 3.1: Straddle Payoff

It is seen that this strategy has very low delta around ATM, so trading ATM straddles can be a way to hedge volatility risk. The ATM straddle will thus play the same role for vega as the swap does for delta.

## 3.5 Short Rate Models

The traditional approach to interest rate modelling was to assume that the entire yield curve was driven by a so-called *short rate*, which is the instantaneous interest rate under continuous compounding. This is a special case of the HJM Framework shown in section 3.3. The idea behind short rate models is fairly simple: Consider the ZCB under continuous compounding where the instantaneous interest rate is given by $r(t)$. We can write the price as

$$P(t,T) = \mathbb{E}_t^{\mathbf{Q}}[D(t,T)] \tag{3.48}$$

where we have defined $D(t,T) := e^{-\int_t^T r(t)\,\mathrm{d}t}$. Given the $\mathbf{Q}$-dynamics of $r(t)$ we can thus find ZCB prices if certain conditions are met. For the one factor models it is assumed that the entire yield curve dynamics are driven by one Wiener process. This makes the models very tractable, but does make the dynamics somewhat primitive. More will be said on this later.

### 3.5.1 Zero Coupon Bond Prices

A class of short rate models that is of particular interest is the *affine models*. These are interesting as there typically exist closed-form formulae for prices on a wide variety of derivatives.

**Proposition 3.8.** *Assume that the short rate follows*

$$\mathrm{d}r(t) = \kappa(\theta - r(t))\,\mathrm{d}t + \sigma\sqrt{\alpha + \beta r(t)}\,\mathrm{d}W(t) \tag{3.49}$$

*where $W(t)$ is a $\mathbf{Q}$-Wiener process and $\kappa > 0, \sigma > 0, \theta, \alpha, \beta$ are constants. If $\beta = 0$ then we must have $\alpha \geq 0$ and vice versa of $\alpha = 0$. Then the price of a Zero Coupon Bond is*

$$P(t,T) = e^{A(t,T) - B(t,T)r(t)} \tag{3.50}$$

*where A and B satisfy the system of Ricatti* ordinary differential equations *(ODEs)*

$$A_t + \frac{1}{2}B^2\sigma^2\alpha - B\kappa\theta = 0 \tag{3.51}$$

$$-B_t + B\kappa + \frac{1}{2}B^2\sigma^2\beta = 1 \tag{3.52}$$

*with boundary condition* $A(T,T) = 0$ *and* $B(T,T) = 0$.

*Proof.* Consider the function $F(t,r) = e^{A(t)-B(t)r}$. We make a qualified guess at the form of $P(t,T) = F(t,r(t))$. Partial derivatives of this function are

$$F_t = F(A_t - B_t r), \qquad F_r = -BF, \qquad F_{rr} = B^2F$$

The terminal condition is $P(T,T) = 1$, so consequently $A(T,T) = 0$ and $B(T,T) = 0$. Using Itô's lemma the dynamics of $F$ can be found as

$$\begin{aligned} \mathrm{d}F &= F_t\,\mathrm{d}t + F_r\,\mathrm{d}r(t) + \frac{1}{2}F_{rr}(\mathrm{d}r(t))^2 \\ &= F(A_t - B_t r(t))\,\mathrm{d}t \\ &\quad + (-BF)\left[\kappa(\theta - r(t))\,\mathrm{d}t + \sigma\sqrt{\alpha + \beta r(t)}\,\mathrm{d}W(t)\right] \\ &\quad + \frac{1}{2}B^2F\sigma^2(\alpha + \beta r(t))\,\mathrm{d}t \end{aligned} \tag{3.53}$$

For the model to be arbitrage-free we must have that the expected return under $\mathbf{Q}$ is $r(t)\,\mathrm{d}t$. This will make the discounted price process a (local) martingale

$$\mathbb{E}_t^{\mathbf{Q}}\left[\frac{\mathrm{d}F}{F}\right] = r(t)\,\mathrm{d}t \qquad \Longleftrightarrow \qquad \frac{\mathbb{E}_t^{\mathbf{Q}}[\mathrm{d}F]}{F\,\mathrm{d}t} - r(t) = 0$$

Combined with the dynamics in equation 3.53 the left hand side of the no arbitrage condition becomes

$$\begin{aligned} \frac{\mathbb{E}_t^{\mathbf{Q}}[\mathrm{d}F]}{F\,\mathrm{d}t} - r(t) &= A_t - B_t r(t) - B\kappa(\theta - r(t)) + \frac{1}{2}B^2\sigma^2\left(\alpha + \beta r(t)\right) - r(t) \\ &= A_t + r(t)\left(-B_t + B\kappa + \frac{1}{2}B^2\sigma^2\beta - 1\right) - B\kappa\theta + \frac{1}{2}B^2\sigma^2\alpha \\ &= \left(A_t + \frac{1}{2}B^2\sigma^2\alpha - B\kappa\theta\right) + r\left(-B_t + B\kappa + \frac{1}{2}B^2\sigma^2\beta - 1\right) \end{aligned}$$

For the non-trivial case of $r(t) \neq 0$ the no arbitrage condition is equivalent to the system of Ricatti ODE's

$$A_t + \frac{1}{2}B^2\sigma^2\alpha - B\kappa\theta = 0$$

$$-B_t + B\kappa + \frac{1}{2}B^2\sigma^2\beta = 1$$

which was the result that we were trying to proof. ■

Solving the Riccati ODEs is by no means trivial. Below the Riccati system is solved for two special cases of the affine model.

**Proposition 3.9.** *Assume that we are in the special case of* $\alpha = 0$ *and* $\beta = 1$ *in Proposition 3.8. This is the* Cox-Ingersoll-Ross *(CIR) model. Then*

$$B(t,T) = \frac{2\left(1 - e^{-\gamma(T-t)}\right)}{(\kappa+\gamma)(1-e^{-\gamma(T-t)}) + 2\gamma e^{-\gamma(T-t)}} \tag{3.54}$$

*and*

$$A(t,T) = \kappa\theta\frac{\gamma+\kappa}{\sigma^2}(T-t) - \frac{2\kappa\theta}{\sigma^2}\log\left(1 + \frac{(\gamma+\kappa)(e^{\gamma(T-t)}-1)}{2\gamma}\right) \tag{3.55}$$

Note that this is also seen in Proposition 10.2.4 in [Andersen and Piterbarg (2010), p. 435], where $c_1 = 0$ and $c_2 = 1$.

*Proof.* The second Ricatti ODE will be solved first as it only depends on $B(t)$[10]. We will solve this for a general $\beta$, so the expression can be used for the Vašíček model by setting $\beta = 0$.

Shorthand notation for $A$ and $B$ will be used by dropping the parenthesis. Consider

$$B = -\frac{2}{\sigma^2\beta}\frac{C_t}{C} \tag{3.56}$$

which transforms the Riccati ODE to a linear second order differential equation, which is considerably easier to solve

$$\frac{2}{\sigma^2\beta}\left(\frac{C_{tt}}{C} - \frac{C_t^2}{C^2}\right) + \left(-\frac{2}{\sigma^2\beta}\frac{C_t}{C}\right)\kappa + \frac{1}{2}\left(-\frac{2}{\sigma^2\beta}\frac{C_t}{C}\right)^2\sigma^2\beta = 1$$

$$\Longleftrightarrow$$

$$\frac{2}{\sigma^2\beta}\left(\frac{C_{tt}}{C} - \frac{C_t^2}{C^2}\right) - \frac{2}{\sigma^2\beta}\frac{C_t}{C}\kappa + \frac{2}{\sigma^2\beta}\frac{C_t^2}{C^2} = 1$$

$$\Longleftrightarrow$$

$$\frac{2}{\sigma^2\beta}C_{tt} - \frac{2}{\sigma^2\beta}C_t\kappa = C$$

$$\Longleftrightarrow$$

$$C_{tt} - \kappa C_t - \frac{\sigma^2\beta}{2}C = 0$$

with boundary conditions $C(T) \neq 0$ and $C'(T) = 0$ such that $B(T) = 0$. The solution to the linear second order differential equation has the general form

$$x^2 - \kappa x - \frac{\sigma^2\beta}{2} = 0 \quad \Longleftrightarrow \quad x = \frac{\kappa \pm \sqrt{\kappa^2 + 2\sigma^2\beta}}{2}$$

[10]This proof is along the same lines as my Hand-In #1 in the course Continuous Time Finance 2 at Copenhagen University.

Note that the solution is always real as $\kappa, \sigma > 0$ and $\beta \geq 0$. Define $\gamma := \sqrt{\kappa^2 + 2\sigma^2\beta}$. The solution can thus be written as

$$C(t) = c_1 e^{\frac{\kappa+\gamma}{2}t} + c_2 e^{\frac{\kappa-\gamma}{2}t} \tag{3.57}$$

where the constants $c_1, c_2$ are found using the boundary condition. The first order derivative is

$$C'(t) = \frac{\kappa+\gamma}{2} c_1 e^{\frac{\kappa+\gamma}{2}t} + \frac{\kappa-\gamma}{2} c_2 e^{\frac{\kappa-\gamma}{2}t} \tag{3.58}$$

so using the boundary condition $C'(T) = 0$ makes it possible to write

$$\frac{\kappa+\gamma}{2} c_1 e^{\frac{\kappa+\gamma}{2}T} + \frac{\kappa-\gamma}{2} c_2 e^{\frac{\kappa-\gamma}{2}T} = 0$$

which has the soluation

$$c_2 = -c_1 \frac{\kappa+\gamma}{\kappa-\gamma} e^{\gamma T}$$

This can be plugged into equation 3.57

$$C(t) = c_1 e^{\frac{\kappa+\gamma}{2}t} - c_1 \frac{\kappa+\gamma}{\kappa-\gamma} e^{\gamma T} e^{\frac{\kappa-\gamma}{2}t} = c_1 \left( e^{\frac{\kappa+\gamma}{2}t} - \frac{\kappa+\gamma}{\kappa-\gamma} e^{\frac{\kappa-\gamma}{2}t+\gamma T} \right) \tag{3.59}$$

And similarly for equation 3.58

$$C'(t) = \frac{\kappa+\gamma}{2} c_1 e^{\frac{\kappa+\gamma}{2}t} - c_1 \frac{\kappa+\gamma}{\kappa-\gamma} e^{\gamma T} \frac{\kappa-\gamma}{2} e^{\frac{\kappa-\gamma}{2}t} = \frac{\kappa+\gamma}{2} c_1 \left( e^{\frac{\kappa+\gamma}{2}t} - e^{\frac{\kappa-\gamma}{2}t+\gamma T} \right) \tag{3.60}$$

These two expressions can then be plugged into equation 3.56

$$\begin{aligned} B(t) &= -\frac{2}{\sigma^2\beta} \frac{\frac{\kappa+\gamma}{2} c_1 \left( e^{\frac{\kappa+\gamma}{2}t} - e^{\frac{\kappa-\gamma}{2}t+\gamma T} \right)}{c_1 \left( e^{\frac{\kappa+\gamma}{2}t} - \frac{\kappa+\gamma}{\kappa-\gamma} e^{\frac{\kappa-\gamma}{2}t+\gamma T} \right)} \\ &= -\frac{\kappa+\gamma}{\sigma^2\beta} \frac{e^{\frac{\kappa+\gamma}{2}t} - e^{\frac{\kappa-\gamma}{2}t+\gamma T}}{e^{\frac{\kappa+\gamma}{2}t} - \frac{\kappa+\gamma}{\kappa-\gamma} e^{\frac{\kappa-\gamma}{2}t+\gamma T}} \end{aligned}$$

The fraction can be reduced by dividing both the numerator and denominator by $e^{\frac{\kappa-\gamma}{2}t+\gamma T}$

$$B(t) = -\frac{\kappa+\gamma}{\sigma^2\beta} \frac{e^{-\gamma(T-t)} - 1}{e^{-\gamma(T-t)} - \frac{\kappa+\gamma}{\kappa-\gamma}} \tag{3.61}$$

Using the expression

$$(\kappa+\gamma)(\kappa-\gamma) = \kappa^2 - \kappa\gamma + \kappa\gamma - \gamma^2 = \kappa^2 - (\kappa^2 + 2\sigma^2\beta) = -2\sigma^2\beta$$

and multiplying the numerator and denominator by $(\kappa - \gamma)$ will reduce equation 3.61 to

$$\begin{aligned} B(t) &= 2\frac{e^{-\gamma(T-t)} - 1}{(\kappa-\gamma)e^{-\gamma(T-t)} - (\kappa+\gamma)} \\ &= 2\frac{1 - e^{-\gamma(T-t)}}{(\kappa+\gamma) - (\kappa-\gamma)e^{-\gamma(T-t)}} \\ &= \frac{2\left(1 - e^{-\gamma(T-t)}\right)}{(\kappa+\gamma)(1-e^{-\gamma(T-t)}) + 2\gamma e^{-\gamma(T-t)}} \end{aligned} \tag{3.62}$$

Going forward $\beta$ will be set to 1. In general a differential equation on the form $\frac{\mathrm{d}A(t)}{\mathrm{d}t} = f(t)$ has the solution $A(t) = \int^t f(s)\,\mathrm{d}s + C$, where $C$ is an arbitrary integration constant. With the boundary condition $A(T) = 0$ this becomes $A(t) = \int_T^t f(s)\,\mathrm{d}s$. We can thus find

$$A(t) = \int_T^t (B(s)\kappa\theta)\,\mathrm{d}s = -\int_t^T (B(s)\kappa\theta)\,\mathrm{d}s = -\kappa\theta\int_t^T B(s)\,\mathrm{d}s \tag{3.63}$$

Calculating the integral

$$\begin{aligned} \int_t^T B(s)\,\mathrm{d}s &= \int_t^T \left(\frac{2\left(1-e^{-\gamma(T-s)}\right)}{(\kappa+\gamma)(1-e^{-\gamma(T-s)}) + 2\gamma e^{-\gamma(T-s)}}\right)\mathrm{d}s \\ &= -2\left[\frac{s}{\gamma-\kappa} + \frac{2\log(\gamma-\kappa+(\gamma+\kappa)e^{\gamma(T-s)})}{\gamma^2-\kappa^2}\right]_t^T \\ &= -2\frac{T-t}{\gamma-\kappa} - \frac{4}{\gamma^2-\kappa^2}\left(\log(\gamma-\kappa+(\gamma+\kappa)) - \log(\gamma-\kappa+(\gamma+\kappa)e^{\gamma(T-t)}\right) \\ &= -2\frac{\gamma+\kappa}{(\gamma-\kappa)(\gamma+\kappa)}(T-t) - \frac{4}{2\sigma^2}\log\left(\frac{2\gamma}{\gamma-\kappa+(\gamma+\kappa)e^{\gamma(T-t)}}\right) \\ &= -2\frac{\gamma+\kappa}{2\sigma^2}(T-t) - \frac{2}{\sigma^2}\log\left(\frac{2\gamma}{(\gamma+\kappa)(e^{\gamma(T-t)}-1)+2\gamma}\right) \\ &= -\frac{\gamma+\kappa}{\sigma^2}(T-t) + \frac{2}{\sigma^2}\log\left(1 + \frac{(\gamma+\kappa)(e^{\gamma(T-t)}-1)}{2\gamma}\right) \end{aligned}$$

and substituting into 3.63 yields

$$A(t) = \kappa\theta\frac{\gamma+\kappa}{\sigma^2}(T-t) - \frac{2\kappa\theta}{\sigma^2}\log\left(1 + \frac{(\gamma+\kappa)(e^{\gamma(T-t)}-1)}{2\gamma}\right) \tag{3.64}$$

We have now proven equations 3.64 and 3.62 where $\gamma = \sqrt{\kappa^2 + 2\sigma^2}$, which are the same as in the proposition. ■

**Proposition 3.10.** *Assume that we are in the special case of $\alpha = 1$ and $\beta = 0$ in Proposition 3.8 (the Vašíček model). Then*

$$B(t,T) = \frac{1 - e^{-\kappa(T-t)}}{\kappa} \tag{3.65}$$

*and*

$$A(t,T) = \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)[(T-t) - B(t,T)] - \frac{\sigma^2 B(t,T)^2}{4\kappa} \tag{3.66}$$

Note that this is the same as in Proposition 10.1.4 in [Andersen and Piterbarg (2010), p. 414].

*Proof.* $B(t)$ in equation 3.62 from the proof to Proposition 3.9 where $\beta = 0$, such that $\gamma = \sqrt{\kappa^2} = \kappa$ gives the $B(t)$ for the Vašíček model

$$B(t,T) = \frac{2(1-e^{-\kappa(T-t)})}{(\kappa+\kappa)(1-e^{-\kappa(T-t)}) + 2\kappa e^{-\kappa(T-t)}} = \frac{2(1-e^{-\kappa(T-t)})}{2\kappa} = \frac{1-e^{-\kappa(T-t)}}{\kappa} \tag{3.67}$$

From Proposition 3.8 the second ODE is given by

$$A_t = B\kappa\theta - \frac{1}{2}B^2\sigma^2 \tag{3.68}$$

$A(t)$ can be found explicitly as

$$A(t) = \int_t^T \frac{1}{2}B^2(s)\sigma^2 - B(s)\kappa\theta \,\mathrm{d}s = \frac{1}{2}\sigma^2 \int_t^T B^2(s)\,\mathrm{d}s - \kappa\theta \int_t^T B(s)\,\mathrm{d}s \tag{3.69}$$

Calculating the integral

$$\begin{aligned} \int_t^T B(s)\,\mathrm{d}s &= \int_t^T \frac{1-e^{-\kappa(T-s}}{\kappa}\,\mathrm{d}s \\ &= \frac{1}{\kappa}\left[s - \frac{e^{-\kappa(T-s)}}{\kappa}\right]_t^T \\ &= \frac{1}{\kappa}(T-t) - \frac{1}{\kappa^2}\left(1-e^{-\kappa(T-t)}\right) \\ &= \frac{1}{\kappa}\left[(T-t) - B(t)\right] \end{aligned} \tag{3.70}$$

And the other integral

$$\int_t^T B^2(s)\,\mathrm{d}s = \int_t^T \frac{(1-e^{-\kappa(T-s)})^2}{\kappa^2}\,\mathrm{d}s = \frac{1}{\kappa^2}\int_t^T 1 + e^{-2\kappa(T-s)} - 2e^{-\kappa(T-s)}\,\mathrm{d}s$$

evaluating the antiderivative

$$\frac{1}{\kappa^2}\left[s + \frac{e^{-2\kappa(T-s)}}{2\kappa} - \frac{2}{\kappa}e^{-\kappa(T-s)}\right]_t^T = \frac{1}{\kappa^2}(T-t) + \frac{1}{2\kappa^3}(1-e^{-2\kappa(T-t)}) - \frac{2}{\kappa^3}(1-e^{-\kappa(T-t)})$$

making it possible to simplify the expression to

$$\int_t^T B^2(s)\,\mathrm{d}s = \frac{1}{\kappa^2}\left[(T-t)-B(t)\right] + \frac{1}{2\kappa^3}(1-e^{-2\kappa(T-t)}) - \frac{1}{\kappa^3}(1-e^{-\kappa(T-t)})$$
$$\iff$$
$$\int_t^T B^2(s)\,\mathrm{d}s = \frac{1}{\kappa^2}\left[(T-t)-B(t)\right] - \frac{1}{2\kappa^3}(1+e^{-2\kappa(T-t)} - 2e^{-\kappa(T-t)})$$

which leads us to the final result

$$\int_t^T B^2(s)\,\mathrm{d}s = \frac{1}{\kappa^2}\left[(T-t)-B(t)\right] - \frac{1}{2\kappa}B^2(t) \tag{3.71}$$

Substituting equations 3.70 and 3.71 into equation 3.69 yields

$$\begin{aligned} A(t) &= \frac{1}{2}\sigma^2\left(\frac{1}{\kappa^2}\left[(T-t)-B(t)\right] - \frac{1}{2\kappa}B^2(t)\right) - \kappa\theta\frac{1}{\kappa}\left[(T-t)-B(t)\right] \\ &= \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)\left[(T-t)-B(t)\right] - \frac{\sigma^2 B^2(t)}{4\kappa} \end{aligned} \tag{3.72}$$

So the proposition is now proved as equations 3.72 and 3.67 has been shown. ■

### 3.5.2 Distribution of the Short-Rate

In Vašíček the short-rate dynamics are given by

$$\mathrm{d}r(t) = \kappa(\theta - r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \tag{3.73}$$

Define $Z(t) := e^{\kappa t}r(t)$, so $Z(0) = r(0)$. Using Itô's lemma to find dynamics

$$\begin{aligned} \mathrm{d}Z(t) &= \kappa e^{\kappa t} r(t)\,\mathrm{d}t + e^{\kappa t}\,\mathrm{d}r(t) \\ &= e^{\kappa t}\left(\kappa r(t)\,\mathrm{d}t + \kappa\theta\,\mathrm{d}t - \kappa r(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)\right) \\ &= e^{\kappa t}\left(\kappa\theta\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)\right) \end{aligned}$$

Substituting $r(t) = e^{-\kappa t}Z(t)$ into the expression yields

$$\begin{aligned} r(t) &= e^{-\kappa t}Z(0) + e^{-\kappa t}\int_0^t e^{\kappa s}\kappa\theta\,\mathrm{d}s + e^{-\kappa t}\int_0^t e^{\kappa s}\sigma\,\mathrm{d}W(s) \\ &= e^{-\kappa t}r(0) + e^{-\kappa t}\kappa\theta\frac{e^{\kappa t}-1}{\kappa} + \sigma\int_0^t e^{-\kappa(t-s)}\,\mathrm{d}W(s) \\ &= \theta + (r(0)-\theta)e^{-\kappa t} + \sigma\int_0^t e^{-\kappa(t-s)}\,\mathrm{d}W(s) \end{aligned} \tag{3.74}$$

From Lemma 4.15 in [Björk (2009), p. 57] it is known that above stochastic integral has mean zero and variance $\sigma^2 \int_0^t e^{-2\kappa(t-s)}\,\mathrm{d}s = \sigma^2 \frac{1-e^{-2\kappa t}}{2\kappa}$. The distribution of the short rate in Vašíček is thus given by

$$\mathbb{E}_0^{\mathbf{Q}}[r(t)] = \theta + (r(0) - \theta)e^{-\kappa t} \tag{3.75}$$

$$\mathrm{Var}_0^{\mathbf{Q}}[r(t)] = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right) \tag{3.76}$$

This distribution will be used in the the Monte-Carlo simulation later on.

In CIR the short rate follows the SDE

$$\mathrm{d}r(t) = \kappa(\theta - r(t))\,\mathrm{d}t + \sigma\sqrt{r(t)}\,\mathrm{d}W(t) \tag{3.77}$$

We have the Feller constraint $2\kappa\theta > \sigma^2$ to make sure that the origin is inaccessible. It can be shown that the conditional density of $r(t)$ follows a noncentral $\chi^2$ distribution with $\nu$ degrees of freedom and non-centrality parameter equal to $\eta$ [Brigo, Morini and Pallavicini (2013), p. 73]. The distribution is specified below

$$f_{r(t)}(x) = c f_{\chi^2(\nu,\eta)}(cx) \tag{3.78}$$

where

$$c = \frac{4\kappa}{\sigma^2(1 - e^{-\kappa t})}, \qquad \nu = \frac{4\kappa\theta}{\sigma^2}, \qquad \eta = cr(0)e^{-\kappa t} \tag{3.79}$$

Values from this distribution can be generated using the inverse transform method

$$r(t) \stackrel{d.}{=} \frac{F^{-1}_{\chi^2(\nu,\eta)}(U)}{c} \tag{3.80}$$

where $U$ is a uniform random variable between 0 and 1. It will later be shown that this is computationally expensive compared to inverse transform sampling from a normal distribution.

### 3.5.3 European Options

First shown by [Jamshidian (1989)], we are able to price options on bonds as well. This is based on the derivations in HJM earlier. Even though Vašíček is not a HJM model as $\theta$ is constant, the price of an option does not depend on the mean level of rates, so we can apply the same techniques here.

**Corollary 3.11 (Options on ZCB in Vašíček).** *Using Proposition 3.5 we can price an option on a ZCB in Vašíček by setting*

$$v = \left(\frac{1 - e^{-\kappa(T_M - T_E)}}{\kappa}\right)^2 \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(T_E - t)}) \tag{3.81}$$

*Proof.* With $d = 1$ and

$$\sigma_P(t,T) = e^{\kappa t}\sigma \int_t^T e^{-\kappa s}\,\mathrm{d}s = \sigma \int_t^T e^{-\kappa(s-t)}\,\mathrm{d}s = \sigma\frac{1 - e^{-\kappa(T-t)}}{\kappa} \tag{3.82}$$

we have

$$
\begin{aligned}
v &= \int_t^{T_E} \left(\sigma_P(u,T_M) - \sigma_P(u,T_E)\right)^2 \mathrm{d}u \\
&= \frac{\sigma^2}{\kappa^2} \int_t^{T_E} \left(e^{-\kappa(T_E-u)} - e^{-\kappa(T_M-u)}\right)^2 \mathrm{d}u \\
&= \frac{\sigma^2}{\kappa^2} \int_t^{T_E} e^{-2\kappa(T_E-u)} + e^{-2\kappa(T_M-u)} - 2e^{-\kappa(T_E+T_M-2u)} \,\mathrm{d}u \\
&= \frac{\sigma^2}{\kappa^2} \left(\frac{1-e^{-2\kappa(T_E-t)}}{2\kappa} + \frac{e^{-2\kappa(T_M-T_E)} - e^{-2\kappa(T_M-t)}}{2\kappa} + \frac{e^{-\kappa(T_E+T_M-2t)} - e^{-\kappa(T_M-T_E)}}{\kappa}\right) \\
&= \frac{\sigma^2}{2\kappa^3}\left(1 - e^{2\kappa(T_E-t)} + e^{-2\kappa(T_M-T_E)} - e^{-2\kappa(T_M-t)} + 2e^{-\kappa(T_E+T_M-2t)} - 2e^{-\kappa(T_M-T_E)}\right)
\end{aligned}
\tag{3.83}
$$

As $\left(1-e^{-\kappa(T_M-T_E)}\right)^2 = 1 + e^{-2\kappa(T_M-T_E)} - 2e^{-\kappa(T_M-T_E)}$ Then

$$
\begin{aligned}
\left(1-e^{-\kappa(T_M-T_E)}\right)^2\left(1-e^{-2\kappa(T_E-t)}\right) = &\left(1 + e^{-2\kappa(T_M-T_E)} - 2e^{-\kappa(T_M-T_E)}\right) \\
&- \left(e^{-2\kappa(T_E-t)} + e^{-2\kappa(T_M-t)} - 2e^{-\kappa(T_M+T_E-2t)}\right)
\end{aligned}
\tag{3.84}
$$

So we can clearly write

$$
v = \left(\frac{1-e^{-\kappa(T_M-T_E)}}{\kappa}\right)^2 \frac{\sigma^2}{2\kappa}(1-e^{-2\kappa(T_E-t)})
\tag{3.85}
$$

So the result is now proven. ■

We note that this is the same as in Proposition 24.9 in [Björk (2009), p. 386] and that the expressions coincides for both the Vašíček and the Hull-White model. See section 4 in [Cox, Ingersoll and Ross (1985)] for the closed-form price of a Zero Coupon Bond call option in the CIR model.

As mentioned before we are not really interested in the price of a zero coupon option, but rather the price of a swaption. We can use the method proposed in [Jamshidian (1989)] to be able to find a semi-closed form expression for the price of a swaption in a short rate model. Consider the time $T_0$ payoff of a payer swaption

$$
V_{P+}(T_0) = \left(\sum_{i^F(t)} \delta_i^F F(T_0;T_{i-1},T_i)P(T_0,T_i) - \sum_{i^K(t)} \delta_i^K K P(T_0,T_i)\right)^+
\tag{3.86}
$$

We have that

$$
\delta_i^F F(T_0;T_{i-1},T_i)P(t,T_i) = \delta_i^F \left(\frac{1}{\delta_i^F}\left(\frac{1}{P(T_0;T_{i-1},T_i)} - 1\right)\right) P(T_0,T_i) = P(T_0,T_{i-1}) - P(T_0,T_i)
\tag{3.87}
$$

so

$$
\sum_{i^F(t)} \delta_i^F F(T_0;T_{i-1},T_i)P(T_0,T_i) = P(T_0,T_0) - P(T_0,T_{N^F}) = 1 - P(T_0,T_{N_L})
\tag{3.88}
$$

Let us use the notation $P_i(r) = e^{A(T_0,T_i)-B(T_0,T_i)r}$. We can then write

$$V_{P+}(T_0) = \left(1 - P_{N_L}(r(T_0)) - \sum_{i^K(t)} \delta_i^K K P_i(r(T_0))\right)^+ \tag{3.89}$$

We note that we can find the interest rate that makes the swap worth zero at time $T_0$ by solving below equation

$$1 = P_{N_L}(r^*) + \sum_{i^K(t)} \delta_i^K K P_i(r^*) \tag{3.90}$$

We can substitute this into equation (3.89)

$$V_{P+}(T_0) = \left(P_{N_L}(r^*) + \sum_{i^K(t)} \delta_i^K K P_i(r^*) - P_{N_L}(r(T_0)) - \sum_{i^K(t)} \delta_i^K K P_i(r(T_0))\right)^+ \tag{3.91}$$

As we pay fixed in the swap, the swaption is valuable when $r(T_0) > r^*$, so we can rewrite

$$\begin{aligned} V_{P+}(T_0) &= \left(P_{N_L}(r^*) - P_{N_L}(r(T_0)) - \sum_{i^K(t)} \delta_i^K K\left(P_i(r^*) - P_i(r(T_0))\right)\right) \mathbb{1}_{\{r(T_0)>r^*\}} \\ &= (P_{N_L}(r^*) - P_{N_L}(r(T_0)))^+ - \sum_{i^K(t)} \delta_i^K K(P_i(r^*) - P_i(r(T_0)))^+ \end{aligned} \tag{3.92}$$

where the last equality follows from monotonicity of the ZCB in terms of the spot rate, which means that

$$P_i(r^*) - P_i(r(T_0)) > 0 \quad \iff \quad r(T_0) > r^* \tag{3.93}$$

We can thus price the swaption as a portfolio of options on ZCB's.

### 3.5.4 Time-Dependent Parameters

One of the problems with above short rate models are that all the parameters are constant, so it will be difficult to calibrate the model to the observed term structure. The approach to make a short rate model that is consistent with observed ZCB prices was the Ho-Lee model. It were set in discrete time, but the continuous version is given by

$$r(t) = r(0) + \varphi(t) + \sigma W(t) \tag{3.94}$$

where $\varphi(t)$ is a deterministic function of time with $\varphi(0) = 0$. Even though that the model is consistent with the observed ZCB, the dynamics are far too simple to be used for any pricing purposes.

In the Hull-White model the SDE for the short rate is specified as

$$\mathrm{d}r(t) = (\theta(t) - \kappa r(t))\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \tag{3.95}$$

This is a special case of Proposition 3.4 when $d = 1$, $\sigma_r(t) = \sigma$ and $\kappa(t) = \kappa$. We can thus match the prices of the observed ZCB by setting

$$
\begin{aligned}
\theta(t) &= \frac{\partial f(0,t)}{\partial t} + \kappa f(0,t) + \sigma^2 \int_0^t e^{-2\kappa(t-u)}\,\mathrm{d}u \\
&= \frac{\partial f(0,t)}{\partial t} + \kappa f(0,t) + \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right)
\end{aligned}
\tag{3.96}
$$

A practical problem with the expression is that it involves $\frac{\partial f(0,t)}{\partial t}$, which can cause issues when the initial yield curve is not smooth. This is for example the case when the curve is found by simple bootstrapping.

We can find the distribution of $r(t)$ in a similar fashion as for the Vašíček model. With $Z(t) = e^{\kappa t} r(t)$ we can use Itô's lemma again

$$
\mathrm{d}Z(t) = e^{\kappa t}\left(\theta(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)\right) \tag{3.97}
$$

so

$$
r(t) = e^{-\kappa t} r(0) + e^{-\kappa t}\int_0^t e^{\kappa s}\theta(s)\,\mathrm{d}s + \sigma\int_0^t e^{-\kappa(t-s)}\,\mathrm{d}W(s) \tag{3.98}
$$

Let

$$
\begin{aligned}
\varphi(t) &= e^{-\kappa t}\int_0^t e^{\kappa s}\theta(s)\,\mathrm{d}s \\
&= e^{-\kappa t}\int_0^t e^{\kappa s}\left(\frac{\partial f(0,s)}{\partial s} + \kappa f(0,s) + \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa s}\right)\right)\mathrm{d}s
\end{aligned}
\tag{3.99}
$$

In general we have $\frac{\partial}{\partial t}\left(e^{\kappa t} f(t)\right) = e^{\kappa t}\left(f'(t) + \kappa f(t)\right)$, so

$$
\begin{aligned}
\varphi(t) &= f(0,t) + e^{-\kappa t}\frac{\sigma^2}{2\kappa}\int_0^t e^{\kappa s} - e^{-\kappa s}\,\mathrm{d}s \\
&= f(0,t) + e^{-\kappa t}\frac{\sigma^2}{2\kappa}\left(\frac{e^{\kappa t}-1}{\kappa} + \frac{e^{-\kappa t}-1}{\kappa}\right) \\
&= f(0,t) + \frac{\sigma^2}{2\kappa^2}\left(1 + e^{-2\kappa t} - 2e^{-\kappa t}\right) \\
&= f(0,t) + \frac{\sigma^2}{2\kappa^2}(1 - e^{-\kappa t})^2 \\
&= f(0,t) + \frac{1}{2}\sigma^2 B(0,t)^2
\end{aligned}
\tag{3.100}
$$

We can now set

$$
r(t) = x(t) + \varphi(t) \tag{3.101}
$$

We have defined the process $x(t)$ with dynamics

$$
\mathrm{d}x(t) = -\kappa x(t)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \tag{3.102}
$$

We set $x(0) = 0$ as $f(0,0) = r(0)$. We now have a process, where the dynamics does not depend on $\frac{\partial f(0,t)}{\partial t}$.

This is a special case of Vašíček, where $\theta = r(0) = 0$, so

$$x(t) = \sigma \int_0^t e^{-\kappa(t-s)}\, \mathrm{d}W(s) \tag{3.103}$$

We can find the forward rate explicitly by integrating the SDE

$$\begin{aligned} f(t,T) &= f(0,T) + \int_0^t \mathrm{d}f(u,T) \\ &= f(0,T) + \int_0^t \mu_f(s,T)\, \mathrm{d}s + \int_0^t \sigma e^{-\kappa(T-u)}\, \mathrm{d}W(u) \end{aligned} \tag{3.104}$$

We have to find $P(t,T) = e^{-\int_t^T f(t,u)\, \mathrm{d}u}$. Using the definition of the forward rate we can write

$$\int_t^T f(0,u)\, \mathrm{d}u = \int_t^T -\frac{\partial \ln P(0,u)}{\partial u}\, \mathrm{d}u = -(\ln P(0,T) - \ln P(0,t)) \tag{3.105}$$

and

$$\int_t^T \int_0^t \mu_f(s,u)\, \mathrm{d}s\, \mathrm{d}u = \frac{1}{2} v \tag{3.106}$$

where $v$ is given in Corollary 3.11 as

$$v = \frac{\sigma^2}{\kappa} B(t,T)^2 (1 - e^{-2\kappa t}) \tag{3.107}$$

and lastly

$$\int_t^T e^{-\kappa(u-t)} x(t)\, \mathrm{d}u = \frac{1 - e^{-\kappa(T-t)}}{\kappa} x(t) \tag{3.108}$$

which leads us to bond reconstitution formula

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\left(-B(t,T)x(t) - \frac{1}{2} v\right) \tag{3.109}$$

Details are in [Andersen and Piterbarg (2010), pp. 416–417] where $\kappa$ and $\sigma$ are also allowed to vary with time.

Similarly we can extend the CIR model to CIR++ by setting

$$r(t) = x(t) + \varphi(t) \tag{3.110}$$

where $x(t)$ is a CIR process with dynamics

$$\kappa(\theta - x(t))\, \mathrm{d}t + \sigma \sqrt{x(t)}\, \mathrm{d}W(t) \tag{3.111}$$

starting in $x(0) > 0$ and $\varphi(t)$ is a deterministic function that makes the model match the initial yield curve by setting

$$\varphi(t) = f(0,t) - \frac{2\kappa\theta(e^{\gamma t}-1)}{2\gamma + (\kappa+\gamma)(e^{\gamma t}-1)} - x(0)\frac{4\gamma^2 e^{\gamma t}}{\left[2\gamma + (\kappa+\gamma)(e^{\gamma t}-1)\right]^2} \tag{3.112}$$

where $\gamma = \sqrt{\kappa^2 + 2\sigma^2}$ as earlier. The last two terms are the forward rate implied by the CIR model.

We will not go into details with the specifics of calibration as this is similar to what has just been shown for the Hull-White model. The reader is referred to [Andersen and Piterbarg (2010), pp. 437–438] for further information.

## 3.6 Gaussian Two-Factor Model

We will denote this model G2++.

$$r(t) = x_1(t) + x_2(t) + \varphi(t) \tag{3.113}$$

where $\varphi(t)$ is a deterministic function of time and

$$\mathrm{d}x_1(t) = -\kappa_1 x_t\,\mathrm{d}t + \sigma_1\,\mathrm{d}W_1(t), \qquad x_1(0) = 0 \tag{3.114}$$

and

$$\mathrm{d}x_2(t) = -\kappa_2 x_t\,\mathrm{d}t + \sigma_2\,\mathrm{d}W_2(t), \qquad x_2(0) = 0 \tag{3.115}$$

The two Wiener processes are correlated

$$\mathrm{d}W_1(t)\,\mathrm{d}W_2(t) = \rho_{1,2}\,\mathrm{d}t \tag{3.116}$$

This can be seen as a two factor version of the Hull-White model although the model usually known as Hull-White two factor has a slightly different parametrisation. The model should provide a better fit to the volatility surface due to the flexibility of three additional parameters.

Define

$$B(t,T) = \begin{pmatrix} B_1(t,T) \\ B_2(t,T) \end{pmatrix}, \qquad B_i(t,T) = \frac{1}{\kappa_i}(1 - e^{-\kappa_i(T-t)}) \tag{3.117}$$

and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{3.118}$$

We can then write the fitting parameter explicitly as

$$\varphi(t) = f(0,t) + \frac{1}{2}B(0,t)^\top \Sigma B(0,t) \tag{3.119}$$

Here we have

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\left(-B(t,T)^\top x(t) - \frac{1}{2}v\right) \tag{3.120}$$

where $v$ is the variance of the integrated interest rate. The reader is referred to [Andersen and Piterbarg (2010), pp. 491–492] for more information. To price swaptions in this model it is also possible to use Jamshidian's decomposition, by fixing $x_1(t)$ at a given level, while solving for $x_2(t)$ and afterward integrating over all possible $x_1(t)$ to find the unconditional expectation.

## 3.7 Discretisation of SDEs

We have implemented an exact simulation scheme for Vašíček as described in equation 2.23. We are simulating the process over one year ($t = 1$) and with 250 steps. We set the seed of the PRNG to 1 and use 10,000 samples. We will use the distribution shown in section 3.5.2. The parameters used are given below

$$\kappa = 0.05, \qquad \theta = 0.03, \qquad \sigma = 0.01, \qquad r(0) = 0.03 \tag{3.121}$$

We have drawn both the discretised distribution and the true distribution

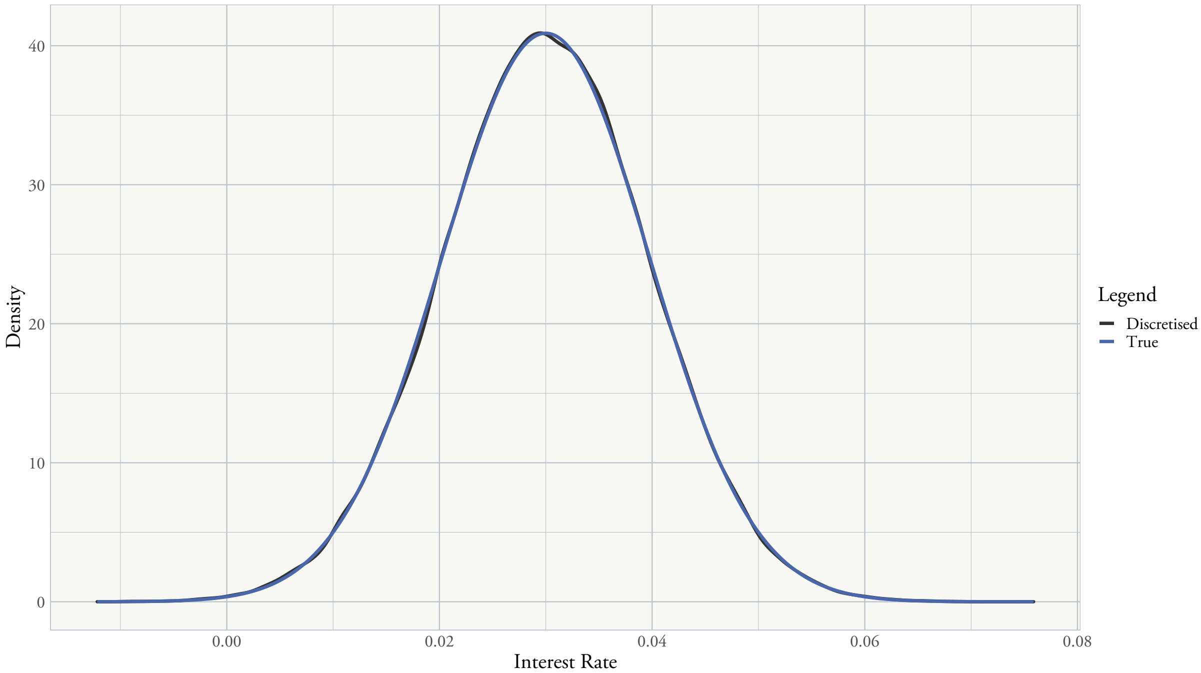



Figure 3.2: Distribution of Discretised Vašíček

We can see that the exact scheme works rather well for Vašíček. It is also of roughly the same speed as a simple Euler scheme.

We will assume that the hazard rate follows a Cox Ingersoll Ross (CIR) process, why it is particularly important that we do not get negative values. The dynamics in CIR are given by[11]

$$dr(t) = \kappa(\theta - r(t))\, dt + \sigma\sqrt{r(t)}\, dW(t) \tag{3.122}$$

where the Feller constraint $2\kappa\theta > \sigma^2$ is satisfied to make sure that the origin is inaccessible. In `QuantLib` the default is to simulate the square root of this process. The idea behind that is that we can use Itô's lemma on

[11] To keep notation in this section consistent we have used $r(t)$. Later we will use $\lambda(t)$ for the hazard rate process to avoid confusion with the interest rate.

$y(t) = f(r(t))$ with $f(x) = \sqrt{x} = x^{\frac{1}{2}}$. We have partial derivatives $f_t = 0$, $f_x = \frac{1}{2}x^{-\frac{1}{2}}$ and $f_{xx} = -\frac{1}{4}x^{-\frac{3}{2}}$, which gives us dynamics of $y(t)$

$$\begin{aligned} \mathrm{d}y(t) &= \frac{1}{2} \cdot [r(t)]^{-\frac{1}{2}} \,\mathrm{d}r(t) + \frac{1}{2} \cdot \left(-\frac{1}{4}\right) \cdot [r(t)]^{-\frac{3}{2}} (\mathrm{d}r(t))^2 \\ &= \frac{1}{2}\left(\kappa\theta\frac{1}{\sqrt{r(t)}} + \kappa\sqrt{r(t)}\right)\mathrm{d}t + \frac{1}{2}\sigma\,\mathrm{d}W(t) - \frac{1}{8}\frac{1}{\sqrt{r(t)}}\sigma^2\,\mathrm{d}t \\ &= \frac{1}{2}\left[\left(\kappa\theta - \frac{\sigma^2}{4}\right)\frac{1}{y(t)} + \kappa y(t)\right]\mathrm{d}t + \frac{\sigma}{2}\,\mathrm{d}W(t) \end{aligned}$$

The advantage of this expression is that the diffusion term does not depend on $y(t)$, which means that we can use the same approach as in Vašíček to simulate from this distribution. There are, however, two drawbacks of this approach:

1. We divide by $y(t)$, so when $y(t)$ gets close to zero then the process 'explodes' meaning that $\mathrm{d}y(t)$ can get infinitely large. To remedy this we can set the drift term to zero, when $y(t) < \varepsilon$. Concretely we have set $\varepsilon = 10^{-3}$, such that the interest rate cannot go below $10^{-6}$. This is by no means a perfect solution, but it will avoid the worst jumps.

2. In each step of the simulation we have to multiply $y(t)$ by itself to find the hazard rate in that point. This might be computationally expensive.

The Quadratic Exponential (QE) discretisation scheme proposed in [Andersen (2007)] will be used as the volatility in Heston follows a CIR process. The approach is based on the notion of moment matching and the reader is referred to the original paper for details. The algorithm will be provided below

1. Calcuate

$$m = \theta + (\hat{r}(t) - \theta)e^{-\kappa\Delta t} \tag{3.123}$$

$$s^2 = \frac{\hat{r}(t)\sigma^2 e^{-\kappa\Delta t}}{\kappa}\left(1 - e^{-\kappa\Delta t}\right) + \frac{\theta\sigma^2}{2\kappa}\left(1 - e^{-\kappa\Delta t}\right)^2 \tag{3.124}$$

$$\psi = \frac{s^2}{m^2} \tag{3.125}$$

2. If $\psi \le \psi_c$

   (a) Calculate

$$b^2 = \frac{2}{\psi} - 1 + \sqrt{\frac{2}{\psi}}\sqrt{\frac{2}{\psi} - 1} \tag{3.126}$$

$$a = \frac{m}{1 + b^2} \tag{3.127}$$

   And set

$$\hat{r}(t + \Delta t) = a(b + X)^2 \tag{3.128}$$

   where $X$ is standard normal.

3. Otherwise if $\psi > \psi_c$

   (a) Calculate

$$p = \frac{\psi - 1}{\psi + 1} \tag{3.129}$$

$$\beta = \frac{1-p}{m} \tag{3.130}$$

And set

$$\hat{r}(t+\Delta t) = \begin{cases} 0 & \text{if } 0 \leq U \leq p \\ \frac{1}{\beta}\ln\left(\frac{1-p}{1-u}\right) & \text{if } p \leq U \leq 1 \end{cases} \tag{3.131}$$

where $U$ is standard uniform.

The first method only works for $\psi \leq 2$ and the second only works for $\psi > 1$, so $\phi_c$ must be between 1 and 2. Following [Andersen (2007)] $\phi_c$ is set to 1.5. Other interesting schemes are also mentioned in the paper, but will not be considered here. It should be noted that the scheme for Heston also involves calculating the stock price, which depend on the CIR process, so that is significantly more complex.

Simple Euler is the scheme proposed in equation 2.23 and the Full Truncation is the same, but where $\hat{r}(t)$ is truncated – it is set to zero if negative.

The density over the terminal distribution for different discretisation schemes are shown below and compared to the true distribution. Parameters are given by

$$\kappa = 0.4, \qquad \theta = 0.026, \qquad \sigma = 0.14, \qquad r(0) = 0.0165 \tag{3.132}$$

And 10,000 paths are used.

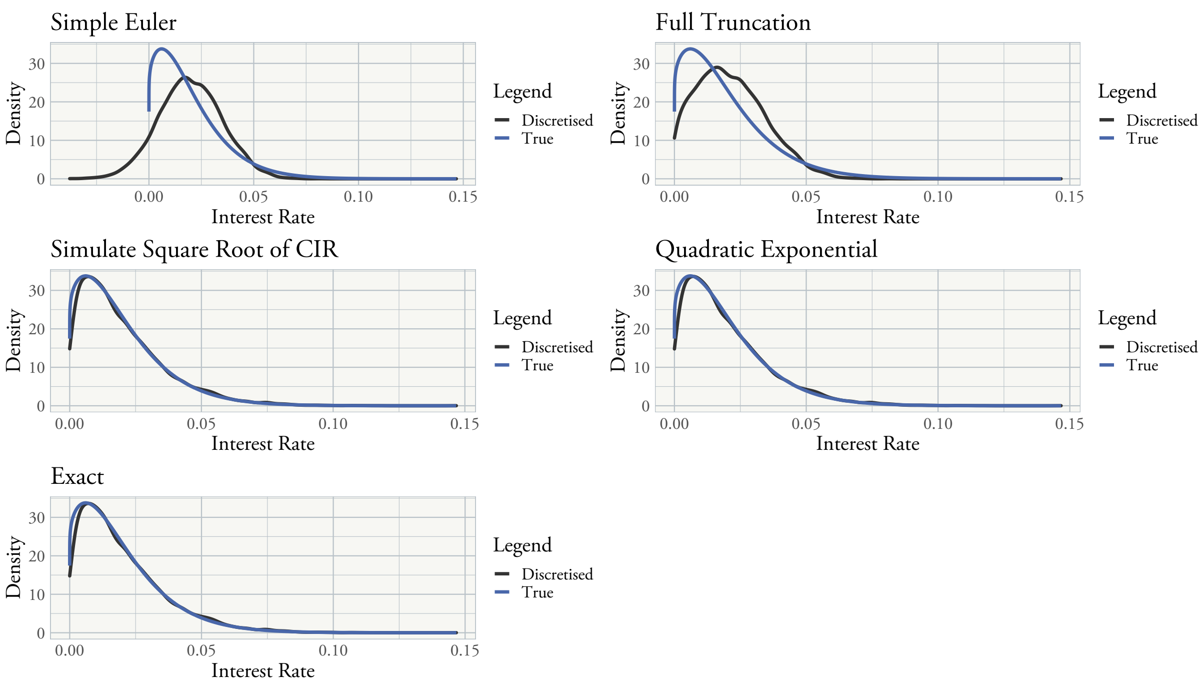



Figure 3.3: Distribution of Discretisation Schemes for CIR

It takes about a second per 10,000 paths for the first four and about 200 seconds for the exact simulation. The exact simulation is over 200 times slower than the discretisation schemes, which makes it impossible to obtain convergence with this approach as we need around a million paths to obtain satisfying convergence. See below for concrete numbers from a run of 10,000 paths

```
Quadratic Exponential: run time = 817.749 ms
Simulate square root of process: run time = 834.093 ms
Full Truncation: run time = 1531.34 ms
Simple Euler discretization: run time = 840.353 ms
Exact simulation: run time = 202555 ms
```

These numbers will of course vary from simulation to simulation, but they give a good indication of the efficiency of the schemes. In this simulation $y(t)$ has only been raised to the power of two at the end of the path and not on every step, so this might overestimate the efficiency of that discretisation scheme. In terms of terminal distribution it is quite difficult to see the advantage of using exact simulation comparing with the schemes that converge. When using the exact distribution one usually do not have to discretise the time grid, but can set the number of steps to one. This would not be feasible as the default event has to be checked at every point on the path. Furthermore there is no closed-form distribution of a $\chi^2$ distribution that is correlated with a Gaussian, so this will also cause issues when introducing dependence between interest rates and intensities, making the exact approach useless. Going forward only the QE scheme will be used and below the convergence is shown for 1,000,000 paths

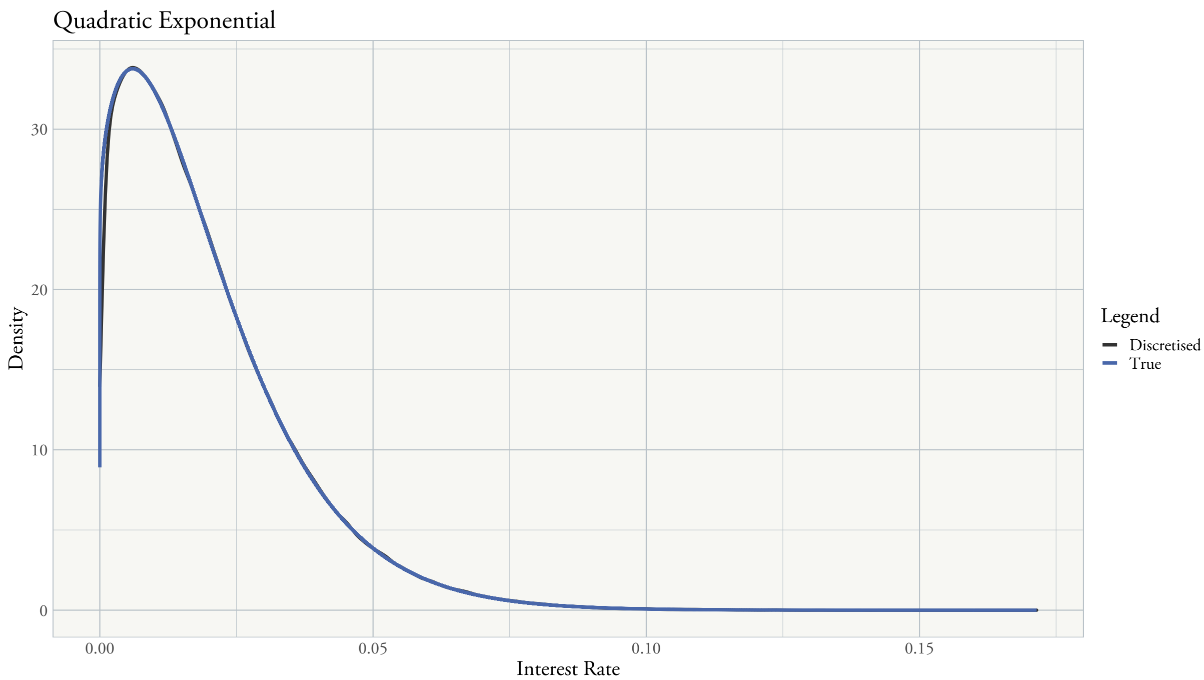



Figure 3.4: Convergence of QE Scheme in CIR

The convergence is deemed satisfying for 1,000,000 paths.

We have now shown discretisation schemes in Vašíček and CIR. The dynamics of CIR++ is just CIR plus a deterministic part and the G2++ is just two Ornsteihn-Uhlenbeck processes with $\theta = 0$, so here we will use the insights from the Vašíček model.

# 4

# Credit Risk Modelling

In this chapter we will introduce the notion of Credit Risk. We will in broad terms base this on the book [Brigo, Morini and Pallavicini (2013)] and will refer the reader to it for further information.

In general there are two types of approaches to the modelling of credit risk: the structural and the intensity approach. The structural approach is concerned around modelling the firm value and considers default as the case where the liabilities exceed the firm value, i.e. default is endogenous to the model. The intensity approach models the probability of default directly and the default is exogenous to the model. We will not go into detail about the structural approach as we find the intensity approach more suitable for pricing purposes, but we will refer to section 3.1 in [Brigo, Morini and Pallavicini (2013)] for an introduction to structural models for credit risk.

## 4.1 Intensity Modelling

For most short-dated derivatives it can be a fairly good assumption that the counterparty is risk-free given that it is of a decent credit quality. But as swaps are usually long-dated, up to 50 years, assuming that the counterparty is risk-free can have a material impact on the value. We will thus account for credit risk in our pricing model. This section follows chapter 4 in [Brigo, Morini and Pallavicini (2013), pp. 65–78].

To this point we have only considered the case, where we enter into a swap with a risk-free counterparty. We will model the default as the first jump in a Poisson process with intensity $\lambda(t)$, which in itself can be a stochastic process. This setup is called a Cox process or a doubly stochastic Poisson process due to the stochastic nature of both the jump and the intensity. We will denote this process $N(t)$ and its natural filtration $\mathscr{H}_t := \sigma(N(s) : 0 \le s \le t)$. We will use the $n$-dimensional stochastic process $X(t)$ to determine default-free market information with natural filtration $\mathscr{F}_t := \sigma(X(s) : 0 \le s \le t)$. We define a new filtration containing all information available at time $t$: $\mathscr{G}_t := \mathscr{F}_t \vee \mathscr{H}_t$. We will in this chapter use the notation $\mathbb{E}_t^{\mathbf{Q}}[X(T)] = \mathbb{E}^{\mathbf{Q}}[X(T)|\mathscr{G}_t]$, so slightly different than last chapter.

**Definition 4.1 (Cox Process).** The Cox process $N(t)$ is a counting process with possibly stochastic intensity $\lambda(t)$. The conditional distribution of changes in the process is Poisson distributed according to

$$\mathbb{P}^{\mathbf{Q}}[N(t) - N(s) = n|\mathscr{F}_t \vee \mathscr{H}_s] = \frac{(\Lambda(s,t))^n}{n!} e^{-\Lambda(s,t)} \tag{4.1}$$

where the cumulative intensity is defined as

$$\Lambda(s,t) := \int_s^t \lambda(u)\,\mathrm{d}u \tag{4.2}$$

Note that this is equivalent of assuming exponentially distributed jump times. In the limit $t \downarrow s$ the probability of a jump is $\lambda(t)\,\mathrm{d}t$, which means that we can write $\mathrm{d}N(t) = \lambda(t)\,\mathrm{d}t$. The intensity is sometimes also called the hazard rate and is the instantaneous probability of default. Obviously we must have $\lambda(t) \ge 0$ a.s. to avoid negative probabilities arising from above expression.

**Proposition 4.2.** *The compensated Cox process*

$$M(t) := N(t) - \int_0^t \lambda(u)\,\mathrm{d}u \tag{4.3}$$

*is a $\mathscr{G}$-martingale.*

*Proof.* Notice that

$$\mathbb{E}^{\mathbf{Q}}[N(t) - N(s)|\mathscr{F}_t \vee \mathscr{H}_s] = \mathbb{E}^{\mathbf{Q}}\left[\int_s^t \mathrm{d}N(u)\middle|\mathscr{F}_t \vee \mathscr{H}_s\right] = \int_s^t \lambda(u)\,\mathrm{d}u$$

so

$$\begin{aligned}
\mathbb{E}^{\mathbf{Q}}[M(t) - M(s)|\mathscr{G}_s] &= \mathbb{E}^{\mathbf{Q}}\left[N(t) - N(s) - \int_s^t \lambda(u)\,\mathrm{d}u\middle|\mathscr{G}_s\right] \\
&= \mathbb{E}^{\mathbf{Q}}\left[\mathbb{E}^{\mathbf{Q}}[N(t) - N(s)|\mathscr{F}_t \vee \mathscr{H}_s] - \int_s^t \lambda(u)\,\mathrm{d}u\middle|\mathscr{G}_s\right] \\
&= 0
\end{aligned}$$

where the second equality is due to the tower property. ■

**Proposition 4.3 (Default Time in Cox).** *The default time $\tau$ will be defined as the first jump in a Cox process. This is equivalent to the first time the cumulative intensity exceeds an independent exponential variable with mean 1 denoted by $\xi$*

$$\tau = \inf\{t : \xi \leq \Lambda(0,t)\} \tag{4.4}$$

The default time is a stopping time as it is measurable with respect to $\mathscr{H}_t$, i.e. we can observe the Cox process to see whether the event $\tau < t$ has happened.

*Proof.* Define

$$\tau := \inf\{t : N(t) \geq 1\} \tag{4.5}$$

The event $\tau > t$ is equivalent with the event $N(t) = 0$ by definition. Using equation 4.1 we can write

$$\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{N(t)=0\}}\middle|\mathscr{F}_t \vee \mathscr{H}_0\right] = e^{-\Lambda(0,t)}$$

Using this can help finding the survival probability

$$\begin{aligned}
\mathbb{P}^{\mathbf{Q}}[N(t) \geq 1|\mathscr{G}_0] &= 1 - \mathbb{P}^{\mathbf{Q}}[N(t) < 1|\mathscr{G}_0] \\
&= 1 - \mathbb{P}^{\mathbf{Q}}[N(t) = 0|\mathscr{G}_0] \\
&= 1 - \mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{N(t)=0\}}\middle|\mathscr{G}_0\right] \\
&= 1 - \mathbb{E}^{\mathbf{Q}}\left[\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{N(t)=0\}}\middle|\mathscr{F}_t \vee \mathscr{H}_0\right]\middle|\mathscr{G}_0\right] \\
&= 1 - \mathbb{E}^{\mathbf{Q}}\left[e^{-\Lambda(0,t)}\middle|\mathscr{G}_0\right]
\end{aligned}$$

Define $\xi$ as a standard exponential variable (mean 1). Using the cumulative distribution function we can write

$$\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{\xi\leq\Lambda(0,t)\}}\middle|\mathscr{F}_t\vee\mathscr{H}_0\right]=1-e^{-\Lambda(0,t)} \tag{4.6}$$

which makes it possible to write

$$\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{N(t)\geq1\}}\middle|\mathscr{G}_0\right]=\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{\xi\leq\Lambda(0,t)\}}\middle|\mathscr{G}_0\right] \tag{4.7}$$

As the events are equal in probability we can redefine $\tau$ as being the first time the cumulative intensity exceeds the value of an independent standard exponential random variable. We can think of the standard exponential random variable as the exogenous default barrier. The default time can be written as

$$\tau=\inf\{t:\xi\leq\Lambda(t)\} \tag{4.8}$$

which was the result to be proven. This also follows from the fact that jump times for the Poisson process are exponentially distributed. ■

**Corollary 4.4 (Survival Probability in Cox).** *The survival probability in a Cox setup is given by*

$$\mathbb{P}^{\mathbf{Q}}[\tau>T|\mathscr{G}_t]=\mathbb{E}^{\mathbf{Q}}\left[e^{-\Lambda(t,T)}\middle|\mathscr{G}_t\right]=\mathbb{1}_{\{\tau>t\}}\mathbb{E}^{\mathbf{Q}}\left[e^{-\Lambda(t,T)}\middle|\mathscr{F}_t\right] \tag{4.9}$$

If we condition on the market price filtration $\mathscr{F}_t$ that does not contain information on whether default has occurred, we have to condition on this explicitly, but it is not an issue as it is known at time $t$. When we use the full filtration $\mathscr{G}_t$ that includes default monitoring this is not needed.

*Proof.* Using results shown in the proof to proposition 4.3 we can write

$$\begin{aligned}\mathbb{P}^{\mathbf{Q}}[\tau>T|\mathscr{G}_t]&=\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{\tau>T\}}\middle|\mathscr{G}_t\right]\\&=\mathbb{E}^{\mathbf{Q}}\left[\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{\tau>T\}}\middle|\mathscr{F}_T\vee\mathscr{H}_t\right]\middle|\mathscr{G}_t\right]\\&=\mathbb{E}^{\mathbf{Q}}\left[e^{-\Lambda(t,T)}\middle|\mathscr{G}_t\right]\\&=\mathbb{1}_{\{\tau>t\}}\mathbb{E}^{\mathbf{Q}}\left[e^{-\Lambda(t,T)}\middle|\mathscr{F}_t\right]\end{aligned}$$ ■

To price derivatives in the Cox setup some 'building blocks' are useful. Define $g(t)$ as generic a $\mathscr{F}_t$-measurable payoff function. Then a default risky time $T$-claim can be prices using

$$\begin{aligned}\mathbb{E}^{\mathbf{Q}}\left[D(t,T)g(T)\mathbb{1}_{\{\tau>T\}}\middle|\mathscr{G}_t\right]&=\mathbb{E}^{\mathbf{Q}}\left[D(t,T)g(T)\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{\tau>T\}}\middle|\mathscr{F}_T\vee\mathscr{H}_t\right]\middle|\mathscr{G}_t\right]\\&=\mathbb{E}^{\mathbf{Q}}\left[D(t,T)g(T)e^{-\Lambda(t,T)}\middle|\mathscr{F}_t\right]\end{aligned}$$

Note that $D(t,T)e^{-\Lambda(t,T)}=e^{-\int_t^T r(u)+\lambda(u)\,\mathrm{d}u}$, which gives $\lambda$ the interpretation of a credit spread. To price a default risky ZCB the risk-free rate is replaced by the risk-free rate plus the default intensity. Now we turn to pricing a continuous stream of payments along the same lines as above

$$\begin{aligned}\mathbb{E}^{\mathbf{Q}}\left[\int_t^T D(t,u)g(u)\mathbb{1}_{\{\tau>u\}}\middle|\mathscr{G}_t\right] &= \mathbb{E}^{\mathbf{Q}}\left[\int_t^T D(t,u)g(u)\mathbb{E}^{\mathbf{Q}}\left[\mathbb{1}_{\{\tau>u\}}\middle|\mathscr{G}_u \vee \mathscr{H}_t\right]\middle|\mathscr{G}_t\right] \\ &= \mathbb{E}^{\mathbf{Q}}\left[\int_t^T D(t,u)g(u)e^{-\Lambda(t,u)}\middle|\mathscr{F}_t\right]\end{aligned}$$

And finally a claim at the time of default can be priced using

$$\begin{aligned}\mathbb{E}^{\mathbf{Q}}\left[D(t,\tau)g(\tau)\mathbb{1}_{\{t<\tau\leq T\}}\middle|\mathscr{G}_t\right] &= \mathbb{E}^{\mathbf{Q}}\left[\mathbb{E}^{\mathbf{Q}}\left[D(t,\tau)g(\tau)\mathbb{1}_{\{t<\tau\leq T\}}\middle|\mathscr{F}_\tau \vee \mathscr{H}_t\right]\middle|\mathscr{G}_t\right] \\ &= \mathbb{1}_{\{\tau>t\}}\mathbb{E}^{\mathbf{Q}}\left[\int_t^T D(t,u)g(u)\lambda(u)e^{-\Lambda(t,u)}\,\mathrm{d}u\middle|\mathscr{F}_t\right]\end{aligned}$$

which follows from the conditional density of the Cox process

$$\frac{\partial\mathbb{P}^{\mathbf{Q}}\left[\tau \leq u|\mathscr{F}_u \vee \mathscr{H}_t\right]}{\partial u} = \lambda(u)e^{-\Lambda(t,u)} \tag{4.10}$$

and from the fact that $\mathbb{1}_{\{\tau>t\}}$ is $\mathscr{G}_t$-measurable.

## 4.2 Credit Default Swaps

As mentioned earlier, it is important to that the models we use are consistent with market data. For credit risk the main derivative for calibration is the *Credit Default Swaps* (CDSs). We could also calibrate to bonds, but CDSs are more standard meaning that they do not differ that much from entity to entity.

The CDS with running coupon is the exchange of a stream of fixed payments for a payment contingent on the default of the *reference entity*[12]. The fixed coupon will be denoted by $c$ and the payment schedule will be denoted[13]

$$i^c(t) := \{i : t < T_i \leq T_{N_c}\} \tag{4.11}$$

It will be assumed that the protection buyer has protection from time $T_0$ to time $T_{N_c}$ and that protection amounts to $1 - \textsc{Rec} = \textsc{Lgd}$ of notional. CDSs can have both physical settlement, where the underlying bond is delivered to the protection buyer, and cash settlement, where the amount is settled in cash upon default. For simplicity it is assumed that the contract is cash-settled. The coverage will be denoted by

$$\delta_i^c := \delta^c(T_{i-1}, T_i) \tag{4.12}$$

If the reference entity defaults within the period $(T_{i-1}, T_i]$, then the interest accrued between $T_{i-1}$ and $\tau$ needs to be payed. We will define $T_-(\tau)$ as the last payment preceding default:

$$T_-(\tau) := \sup\{T_i : T_i \leq \tau\} \tag{4.13}$$

[12] Strictly speaking the payment is contingent on a *credit event*, which can also mean restructuring of debt and other types of technical default. The types of credit events are described in [Brigo, Morini and Pallavicini (2013), pp. 31–32].

[13] The reason for not using $K$ here is to avoid confusion by using the same notation for different things.

The price of a buyer CDS, which is buying protection and thus paying the premium, will be denoted by $V_B(t)$ and a seller CDS, which is selling protection and thus receiving the premium, will be denoted by $V_S(t)$. Note that a buyer CDS can be thought of as a payer IRS where the floating leg is replaced by a protection leg and that a seller CDS has a similar interpretation. The price of a CDS can be written as

$$\begin{aligned} V_B(t) &= \mathbb{E}_t^{\mathbf{Q}}\left[\textsc{Lgd} D(t,\tau)\mathbb{1}_{\{T_0<\tau\le T_{N_c}\}}\right] \\ &\quad - \mathbb{E}_t^{\mathbf{Q}}\left[c\delta^c(T_-(\tau),\tau)D(t,\tau)\mathbb{1}_{\{T_0<\tau\le T_{N_c}\}} + \sum_{i^c(t)} c\delta_i^c D(t,T_i)\mathbb{1}_{\{\tau>T_i\}}\right] \\ &= \textsc{Lgd}\cdot\mathbb{1}_{\{\tau>t\}}\mathbb{E}^{\mathbf{Q}}\left[\int_{T_0}^{T_{N_c}} D(t,u)\lambda(u)e^{-\Lambda(t,u)}\,\mathrm{d}u\middle|\mathscr{F}_t\right] \\ &\quad - c\cdot\mathbb{1}_{\{\tau>t\}}\mathbb{E}^{\mathbf{Q}}\left[\int_{T_0}^{T_{N_c}} D(t,u)\delta^c(T_-(u),u)\lambda(u)e^{-\Lambda(t,u)}\,\mathrm{d}u\middle|\mathscr{F}_t\right] \\ &\quad - c\cdot\mathbb{1}_{\{\tau>t\}}\sum_{i^c(t)}\mathbb{E}^{\mathbf{Q}}\left[D(t,T_i)\delta_i^c e^{-\Lambda(t,T_i)}\middle|\mathscr{F}_t\right] \end{aligned} \tag{4.14}$$

where $\frac{\mathrm{d}\mathbb{P}_t^{\mathbf{Q}}[\tau>s]}{\mathrm{d}s} = -\lambda(s)e^{-\int_t^s\lambda(u)\,\mathrm{d}u}\,\mathrm{d}s$. The pricing consists of three parts: protection, accrual and premium. Assume independence between interest rates and intensities[14] then the mid-point approximation[15] can be used to simplify the formula significantly. Here it is assumed that the reference entity can only default in the middle of a coupon period. Define $T_{i-\frac{1}{2}} := \frac{T_i - T_{i-1}}{2}$. This gives rise to the formula

$$\begin{aligned} V_B(t) &\approx \sum_{i^c(t)} \textsc{Lgd} P(t,T_{i-\frac{1}{2}})\mathbb{P}_t^{\mathbf{Q}}[T_{i-1}<\tau\le T_i] \\ &\quad - c\sum_{i^c(t)}\delta^c(T_{i-1},T_{i-\frac{1}{2}})P(t,T_{i-\frac{1}{2}})\mathbb{P}_t^{\mathbf{Q}}[T_{i-1}<\tau\le T_i] \\ &\quad - c\sum_{i^c(t)}\delta_i^c P(t,T_i)\mathbb{P}_t^{\mathbf{Q}}[\tau>T_i] \end{aligned} \tag{4.15}$$

The default probabilities can be simplified to

$$\begin{aligned} \mathbb{P}_t^{\mathbf{Q}}[T_{i-1}<\tau\le T_i] &= \mathbb{P}_t^{\mathbf{Q}}[\tau\le T_i] - \mathbb{P}_t^{\mathbf{Q}}[\tau\le T_{i-1}] = 1 - \mathbb{P}_t^{\mathbf{Q}}[\tau>T_i] - (1-\mathbb{P}_t^{\mathbf{Q}}[\tau>T_{i-1}]) \\ &= \mathbb{P}_t^{\mathbf{Q}}[\tau>T_{i-1}] - \mathbb{P}_t^{\mathbf{Q}}[\tau>T_i] \end{aligned} \tag{4.16}$$

So the approximation can be written as

$$\begin{aligned} V_B(t) &\approx \sum_{i^c(t)}\left(\textsc{Lgd} - c\delta_i^c(T_{i-1},T_{i-\frac{1}{2}})\right)P(t,T_{i-\frac{1}{2}})\left(\mathbb{P}_t^{\mathbf{Q}}[\tau>T_{i-1}] - \mathbb{P}_t^{\mathbf{Q}}[\tau>T_i]\right) \\ &\quad - c\sum_{i^c(t)}\delta_i^c P(t,T_i)\mathbb{P}_t^{\mathbf{Q}}[\tau>T_i] \end{aligned} \tag{4.17}$$

[14] Correlation between interest rates and default intensities is negligible in valuation of CDS, so this can be disregarded even though that this is not consistent with wrong way risk [Brigo, Morini and Pallavicini (2013), p. 125].

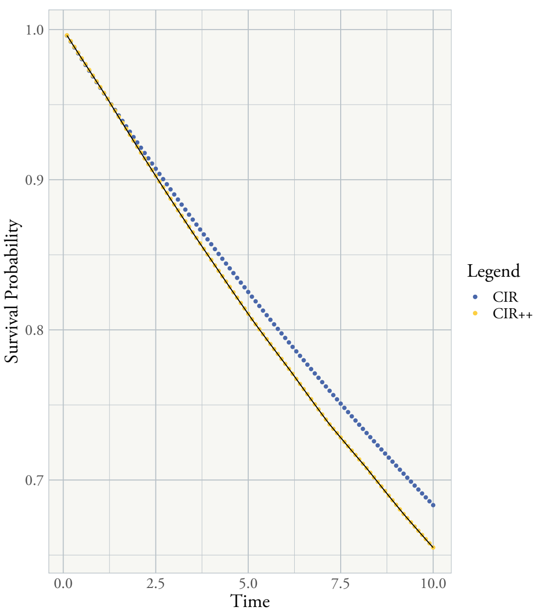
[15] The mid-point approximation is described in detail here. Another approximation worth mentioning is the ISDA Model, which is described here.

As mentioned earlier $\mathbb{P}_t^{\mathbf{Q}}[\tau > T]$ in a Cox setting is equivalent to $P(t,T)$ in a short rate model where $r$ plays the role of $\lambda$. In general the CIR model will be used to model hazard rates. If the Feller constraint is satisfied, then positive hazard rates are ensured. The CIR model is very tractable as a closed-form expression for ZCBs exists, which is essentially a survival probability, when the hazard rate is modelled.

To price a CDS under independence between interest rates and intensities only a term structure for interest rates and a term structure for hazard rates are needed. This also means that we can find the market implied survival probabilities from CDS spreads. Our approach is to first generate a term structure for hazard rates that exactly produces observed prices for CDSs when using the mid-point approximation. Then given that curve the CIR model is calibrated to at-the-money (ATM) *Credit Default Swap Option* (CDSO) prices. CDSOs are also sometimes called Credit Default Swaptions. To calibrate our intensity model a CDSO pricing model is needed. Let $V_{B+}$ be the option to enter a buyer CDS at time $T_0$. The price is given by

$$V_{B+}(t) = \mathbb{E}_t^{\mathbf{Q}}[D(t,T_0)(V_B(T_0))^+] \tag{4.18}$$

This expectation is not trivial to evaluate, but can be simplified by working under the $T_0$-forward measure due to the assumption of independence between interest rates and intensities. According to [Brigo and Alfonsi (2004)] it is possible to do the Jamshidian decomposition on this derivative as well. There is no function in `QuantLib` to price CDSOs in the CIR model, so I have implemented a Monte Carlo simulation. It was tested using input from Table 12.7 in [Brigo, Morini and Pallavicini (2013), p. 290], but it was very slow and it did not seem to work as intended. For some model values the implied volatility could not be found, which suggest that the prices produced by the model are arbitrageable. This is for example the case if the price of the option is lower than the intrinsic value. It would probably have been more efficient to implement the semi closed-form formula using Jamshidian's decomposition. However, recent CDSO data could not be obtained, so the function is of little use. The implementation will be described briefly: A Monte Carlo scheme simulating $\lambda(T_0)$ from $\lambda(t)$ in 'one go', i.e. using the exact distribution of the intensity under CIR, which has been shown earlier. After generating $\lambda(T_0)$ the survival probability curve implied by the CIR model can be found and then the price of the underlying CDS is found using the mid-point approximation. If the value is negative it will be set to zero. When the number of samples gets large this should approach the true value of the CDSO per the central limit theorem. The yield curve is assumed to be deterministic based on the current zero curve. Calibration to the data is shown below



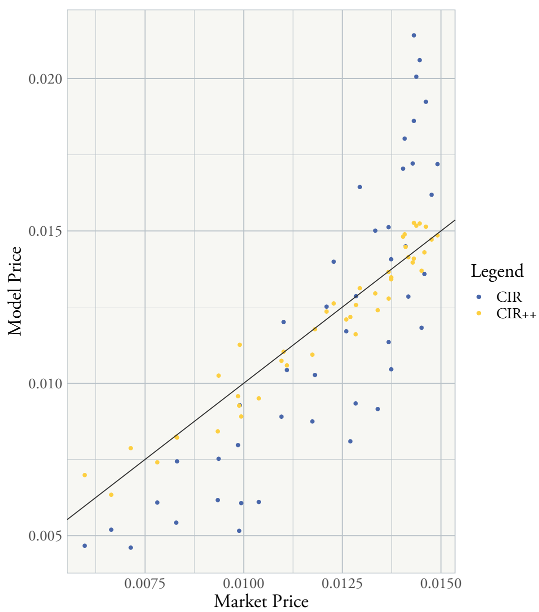




Figure 4.1: Calibration of CIR to CDS's and CDS Options.

Data source is table 12.4 and 12.7 in [Brigo, Morini and Pallavicini (2013), pp. 289–290].

It should be stated that the CIR model is not fitted to the CDS spreads, but only the CDSO prices, while the CIR++ model is consistent with the default probability term structure implied by the CDS spreads. From the right hand graph it seems like the CIR++ fits the prices rather well, so it is actually surprising that there is issues with finding implied volatility for the model prices. The relative error for the CIR++ model is 5% on average, so this does not seem like a large error. It is seen that the CIR model has far too few degrees of freedom to fit the volatility surface, which is as expected.

In terms of risk management the approach shown in section 3.2.2 for swaps and section 3.4.2 for swaptions can be applied to CDSs and CDSOs as well. The model rates are now default intensities, but otherwise it is very similar.

## 4.3 Credit and Debit Value Adjustment

*Credit Value Adjustment* (CVA) is, as the name suggest, the difference between entering the derivative with a default risk-free counterparty and default risky counterparty. When the counterparty can default, the dealer must be compensated for the possibility that it will not be payed back in full in the case of a positive value at default. Similarly the *Debit Value Adjustment* (DVA) is CVA seen from the counterparty's point of view. It is the compensation for the possibility of the dealer defaulting on its obligations towards the counterparty.

Accounting rules (IFRS 13) state that CVA and DVA must be calculated, so the values in the financial report account for credit risk. DVA is however somewhat controversial. In terms of regulatory capital DVA is not included, while CVA is. As DVA increases when the credit quality of the dealer deteriorates, a worsening credit quality would increase equity of the dealer, which is undesirable from a regulatory point of view. There has been examples of banks reporting an enormous positive mark-to-market due to deterioration of their own credit quality. This seems paradoxical as the only way to realise the gain is to default. Hedging can obviously not be direct as the seller cannot coincide with the reference entity in CDSs. To reduce DVA risk a bank would typically sell CDSs on a basket of similar companies, but this only hedges the spread risk and not the jump to default risk of DVA [Brigo, Morini and Pallavicini (2013), pp. 253–255]. This is some undesirable features of DVA, but from a pricing point of view we have to account for DVA – otherwise the parties would not agree on the price.

The swap dealer (bank) will be denoted as $B$ and the counterparty as $C$. If nothing else is stated we will look at the price from the bank's point of view. It will be assumed throughout that we are standing at time $t$ and are pricing a (possibly forward-starting) portfolio with default risk-free price process $\Pi(t)$. Typically the portfolio is determined by the *netting set* defined in the ISDA Master Agreement. At the time of default the value of the trades within the netting set is calculated and the net amount is settled. This can reduce credit risk significantly if trades in opposite directions are traded with the same counterparty. $\Pi(t)$ is the $\mathbf{Q}$-expectation to all the cash flows between $t$ and $T$ deflated by the bank account. Mathematically this can represent this as

$$\Pi(t) := \mathbb{E}_t^{\mathbf{Q}}\left[\int_t^T D(t,s)\,\mathrm{d}PP(s)\right] \tag{4.19}$$

where we have defined $D(t,T) = e^{-\int_t^T r(s)\,\mathrm{d}s}$ as earlier. $PP$ is the cumulative promised payments such that $\int_t^T \mathrm{d}PP(s)$ is the sum of all cash flows from $t$ to $T$ (undiscounted). The deterministic process $PP$ is defined only for intuition and will not be used going forward. Here $\Pi(t)$ is defined slightly different than in [Brigo, Morini and Pallavicini (2013)] as they define it without the expectation. In this notation the Net Present Value (NPV) of the portfolio is $\Pi(t)$. We note that $\Pi(t)$ is the value without any credit risk and we can think of this as the value of the portfolio had it been traded with a hypothetical default risk-free counterparty.

Apart from the default intensity explained earlier, another important aspect of pricing is the *recovery*, e.g. in case of default, what amount is recovered which will be denoted $\textsc{Rec}$. It will be assumed that a share of the market value at the time of default is recovered, so we can model the *Loss Given Default* (LGD) as a fraction of the *Exposure At Default* (EAD). We will in general assume that the LGD is constant even though that this is not realistic[16].

### 4.3.1 Unilateral CVA

At first assume only the counterparty ($C$) can default, which is equivalent of assuming $\tau^C < \tau^B$ almost surely. Given default before time $T$ the bank ($B$) receive following values (receiving a negative value means paying)

| | $t$ | $\tau^C$ |
|---|---|---|
| $\Pi(\tau^C) > 0$ | $\Pi(t, \tau^C)$ | $\textsc{Rec}^C \Pi(\tau^C, T)$ |
| $\Pi(\tau^C) < 0$ | $\Pi(t, \tau^C)$ | $\Pi(\tau^C, T)$ |

Table 4.1: Values of Possible Outcomes Given Early Default

The column names reflect at which point in time the values are related to and the row names reflect whether or not the counterparty owes the bank money at the time of default. The notation $\min(x, 0) = x^-$ and $\max(x, 0) = x^+$ is used. In the case of default before the end of the contract the bank receives

$$\Pi(t, \tau^C) + D(t, \tau^C)\left(\textsc{Rec}^C[\Pi(\tau^C, T)]^+ + [\Pi(\tau^C, T)]^-\right) \tag{4.20}$$

where $\textsc{Rec}^C = 1 - \textsc{Lgd}^C$. As $x = x^- + x^+$ it can be simplified to

$$\Pi(t, \tau) + D(t, \tau^C)\Pi(\tau^C, T) - D(t, \tau)\textsc{Lgd}^C[\Pi(\tau^C)]^+ \tag{4.21}$$

It should be obvious that the first two terms are just the value of the contract at time $t$. It includes all cash flows from $t$ to $T$ discounted back to time $t$. In the case of early default the bank thus receives

$$\Pi(t, T) - D(t, \tau^C)\textsc{Lgd}^C[\Pi(\tau^C)]^+ \tag{4.22}$$

We can now write the payoff of entering the contract with a default risky counterparty as

$$\Pi(t, T)\mathbb{1}_{\{\tau^C > T\}} + \left(\Pi(t, T) - D(t, \tau)\textsc{Lgd}^C[\Pi(\tau^C)]^+\right)\mathbb{1}_{\{t < \tau^C \le T\}} \tag{4.23}$$

which can be rewritten to

$$\Pi(t, T) - D(t, \tau^C)\textsc{Lgd}^C[\Pi(\tau^C)]^+\mathbb{1}_{\{t < \tau^C \le T\}} \tag{4.24}$$

as $\Pi(t, T) = \Pi(t, T)\mathbb{1}_{\{\tau^C > T\}} + \Pi(t, T)\mathbb{1}_{\{t < \tau^C \le T\}}$. The default risky price will be denoted by $\tilde{\Pi}(t)$ and can be

[16] The LGD is used as input when calibrating our intensity model to market prices for CDS's, so the model prices will be consistent with market prices, but it can of course still make a difference which level is chosen.

calculated as

$$\begin{aligned}\tilde{\Pi}(t) &= \mathbb{E}_t^{\mathbf{Q}}\Big[\Pi(t,T) - D(t,\tau^C)\textsc{Lgd}^C[\Pi(\tau^C)]^+\mathbb{1}_{\{t<\tau^C\leq T\}}\Big] \\ &= \Pi(t) - \textsc{Lgd}^C\mathbb{E}_t^{\mathbf{Q}}\Big[D(t,\tau^C)[\Pi(\tau^C)]^+\mathbb{1}_{\{t<\tau^C\leq T\}}\Big]\end{aligned} \tag{4.25}$$

where the second equality is due to $\Pi(t) = \Pi(t,T)$ being $\mathscr{G}_t$-measurable. The value to be subtracted from the default risk-free price is the unilateral CVA. One complexity with above expression is that there is two sources of stochasticity: the default time and the term structure of interest rates. This makes it difficult to simplify the expression in general as we need to know the joint density of $\tau$ and $\Pi$ to evaluate the expectation. This is not easy to model as default is a very rare event and different counterparties might have different relationships between market prices on their swap portfolio and the default probability.

**Example 4.5.** Consider a housing cooperative with a variable rate liability. They want to hedge their exposure towards rising interest rate by entering into a payer swap, where the floating leg matches the rate on the liability. Disregarding that accounting treatment might be different for the liability and the swap, changes in the value of the derivative will not have any impact on the economic situation of the company. Rising interest rates would benefit the swap, but hurt the liability, and thus cancelling out any market movements, which is the definition of a hedge. In this case the probability of default will not be affected by the market value of the swap portfolio.

If swaps were only used for hedging, it would be reasonable to assume independence, but swaps are also used for speculation.

**Example 4.6.** Consider a hedge fund that speculates that interest rates will be higher in the future, so they take an outright position in a payer swap. We assume that this is their only position. Then if interest rates decrease, it will decrease the market value of the position and consequently the fund, thus increasing the probability of default. In this case the bank will have *wrong way risk* towards the counterparty as increasing value of the position for the bank also increases credit risk towards the counterparty.

The second example motivates the definition

**Definition 4.7 (Wrong Way Risk).** *Wrong Way Risk* (WWR) is when EAD from the bank's point of view and the probability of default of the counterparty are positively correlated.

*Right Way Risk* (RWR) is the opposite and will typically occur if only a portion of the exposure is hedged. For instance if the housing cooperative in example 4.5 had only hedged 50% of their notional. In the swap the counterparty is positioned for higher rates, but in conjunction with the liability they are still positioned for lower rates. Increasing rates would thus increase the probability of default for the counterparty. As the bank is receiving in the swap the EAD decreases when rates increase. So EAD and the probability of default are negatively correlated, so this is an example of RWR.

### 4.3.2 Unilateral DVA

Now assume that the counterparty is doing the same analysis with the assumption that the bank can default, but the counterparty cannot. It will be used that the counterparty has the opposite side of the trade of the bank, such that $\Pi^C(t) = -\Pi^B(t)$. Under these assumptions the default risky price becomes

$$\tilde{\Pi}^C(t) = \Pi^C(t) - \textsc{Lgd}^B\mathbb{E}_t^{\mathbf{Q}}\Big[D(t,\tau^B)[\Pi^C(\tau^B)]^+\mathbb{1}_{\{t<\tau^B\leq T\}}\Big] \tag{4.26}$$

From the bank's point of view the value is

$$-\tilde{\Pi}^B(t) = -\Pi^B(t) - \text{LGD}^B \mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau^B)[-\Pi^B(\tau^B)]^+ \mathbb{1}_{\{t<\tau^B\leq T\}}\right] \tag{4.27}$$

such that we can write

$$\tilde{\Pi}^B(t) = \Pi^B(t) + \text{LGD}^B \mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau^B)[-\Pi^B(\tau^B)]^+ \mathbb{1}_{\{t<\tau^B\leq T\}}\right] \tag{4.28}$$

The value to be added to the default risk-free price is the unilateral DVA.

### 4.3.3 Bilateral CVA

Typically both the bank and the counterparty can default. Define $\tau := \min(\tau^B, \tau^C)$, so $\tau$ represent the first-to-default time. Using the building blocks from above it should be obvious that we can write the bilateral adjustment as

$$\begin{aligned}\tilde{\Pi}^B(t) = \Pi^B(t) & \\ &- \text{LGD}^C \mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau^C)[\Pi^B(\tau^C)]^+ \mathbb{1}_{\{t<\tau\leq T\}} \mathbb{1}_{\{\tau=\tau^C\}}\right] \\ &+ \text{LGD}^B \mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau^B)[-\Pi^B(\tau^B)]^+ \mathbb{1}_{\{t<\tau\leq T\}} \mathbb{1}_{\{\tau=\tau^B\}}\right]\end{aligned} \tag{4.29}$$

Define the CVA from the bank's point of view as

$$\text{CVA}^B(t) := \text{LGD}^C \mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau)[\Pi^B(\tau)]^+ \mathbb{1}_{\{t<\tau\leq T\}} \mathbb{1}_{\{\tau=\tau^C\}}\right] \tag{4.30}$$

and similarly for DVA

$$\text{DVA}^B(t) := \text{LGD}^B \mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau)[-\Pi^B(\tau)]^+ \mathbb{1}_{\{t<\tau\leq T\}} \mathbb{1}_{\{\tau=\tau^B\}}\right] \tag{4.31}$$

such that the price with credit risk can be written as

$$\tilde{\Pi}^B(t) = \Pi^B(t) - \text{CVA}^B(t) + \text{DVA}^B(t) \tag{4.32}$$

Note that both CVA and DVA is defined as strictly positive values. Loosely speaking, CVA is a compensation to the bank, which is the reason for subtracting this from the price, while DVA is a compensation to the counterparty, why the amount is added to the price from the bank's point of view. As mentioned earlier DVA is controversial, but without it the default risky price would be asymmetric meaning that the two parties would not agree on the price. So from a pricing point of view it is crucial to include DVA.

The Bilateral Credit Value Adjustment (BVA) can be defined as a quantity to be *subtracted* from the default risk-free price or as a quantity to be *added* to the default risk-free price. As [Brigo, Morini and Pallavicini (2013)] points out, this adjustment can be either positive or negative, so it might be clearer, if the latter approach is used such that $\text{BVA}(t) := -\text{CVA}(t) + \text{DVA}(t)$.

**Proposition 4.8 (Bilateral Credit Value Adjustment).** *Assume that both the bank* ($B$) *and the counterparty* ($C$) *can default. Define* $\tau := \min(\tau^B, \tau^C)$ *and the value process of the netting set as* $\Pi(t)$*. The default risky price is then given by*

$$\tilde{\Pi}^B(t) = \Pi^B(t) + \mathrm{BVA}^B(t) \tag{4.33}$$

*where the Bilateral Credit Value Adjustment (BVA) is given by*

$$\mathrm{BVA}^B(t) := -\mathrm{CVA}^B(t) + \mathrm{DVA}^B(t) \tag{4.34}$$

*and with Credit Value Adjustment (CVA) given by*

$$\mathrm{CVA}^B(t) := \mathrm{LGD}^C \mathbb{E}_t^{\mathbf{Q}}\Big[D(t,\tau)[\Pi^B(\tau)]^+ \mathbb{1}_{\{t<\tau\leq T\}} \mathbb{1}_{\{\tau=\tau^C\}}\Big] \tag{4.35}$$

*and with Debit Value Adjustment (DVA) given by*

$$\mathrm{DVA}^B(t) := \mathrm{LGD}^B \mathbb{E}_t^{\mathbf{Q}}\Big[D(t,\tau)[-\Pi^B(\tau)]^+ \mathbb{1}_{\{t<\tau\leq T\}} \mathbb{1}_{\{\tau=\tau^B\}}\Big] \tag{4.36}$$

*Proof.* The arguments made throughout this section is considered as proof for the proposition. The bottom-up approach presented here is in this case deemed more natural than the proposition-proof approach, which is typically used in this thesis. ∎

One slight inconsistency with above approach is that we have assumed that at the time of default, the value of the portfolio traded with a hypothetical risk-free counterparty is considered. This is called risk-free close-out and it is assumed for simplicity. At default we are thus not accounting for the credit risk of the surviving party. If this was accounted for, the valuation would be more continuous. Close-out is described in detail in chapter 14 of [Brigo, Morini and Pallavicini (2013), pp. 319–330].

### 4.3.4 Approximation

Consider a discretisation of time: Let $T_0$ denote the beginning of the portfolio and $T_n$ the longest maturity of the portfolio. We have $n$ intervals of an arbitrary length – not necessarily equidistant even though this is usually assumed for simplicity. The approximation assuming that default can only occur at the end of the intervals can for CVA be written as

$$\begin{aligned}
\mathrm{CVA}^B(t) &= \mathrm{LGD}^C \sum_{j=1}^{n} \mathbb{E}_t^{\mathbf{Q}}\Big[D(t,\tau)[\Pi^B(\tau)]^+ \mathbb{1}_{\{T_{j-1}<\tau\leq T_j\}} \mathbb{1}_{\{\tau=\tau^C\}}\Big] \\
&\approx \mathrm{LGD}^C \sum_{j=1}^{n} \mathbb{E}_t^{\mathbf{Q}}\Big[D(t,T_j)[\Pi^B(T_j)]^+ \mathbb{1}_{\{T_{j-1}<\tau\leq T_j\}} \mathbb{1}_{\{\tau=\tau^C\}}\Big]
\end{aligned} \tag{4.37}$$

and similarly for DVA

$$\begin{aligned}
\mathrm{DVA}^B(t) &= \mathrm{LGD}^B \sum_{j=1}^{n} \mathbb{E}_t^{\mathbf{Q}}\Big[D(t,\tau)[-\Pi^B(\tau)]^+ \mathbb{1}_{\{T_{j-1}<\tau\leq T_j\}} \mathbb{1}_{\{\tau=\tau^B\}}\Big] \\
&\approx \mathrm{LGD}^B \sum_{j=1}^{n} \mathbb{E}_t^{\mathbf{Q}}\Big[D(t,T_j)[-\Pi^B(T_j)]^+ \mathbb{1}_{\{T_{j-1}<\tau\leq T_j\}} \mathbb{1}_{\{\tau=\tau^B\}}\Big]
\end{aligned} \tag{4.38}$$

This is similar to the mid-point approach to pricing CDS's, but here accrued interest is part of the portfolio value at the time of default. Assuming independence between $\Pi$ and $\tau$ the expression for CVA can be simplified further

$$\text{CVA}^B(t) \approx \text{LGD}^C \sum_{j=1}^{n} \mathbb{E}_t^{\mathbf{Q}}\left[D(t,T_j)[\Pi^B(T_j)]^+\right]\mathbb{P}_t^{\mathbf{Q}}\left[T_{j-1} < \tau \leq T_j, \tau = \tau^C\right] \tag{4.39}$$

and a similar expression is obtained for DVA

$$\text{DVA}^B(t) \approx \text{LGD}^B \sum_{j=1}^{n} \mathbb{E}_t^{\mathbf{Q}}\left[D(t,T_j)[-\Pi^B(T_j)]^+\right]\mathbb{P}_t^{\mathbf{Q}}\left[T_{j-1} < \tau \leq T_j, \tau = \tau^B\right] \tag{4.40}$$

Both CVA and DVA are options on the residual portfolio value. To determine CVA and DVA only a model for (basket) options and a model for survival probabilities is needed. It is still not trivial to calculate CVA and DVA and the approximation is based on some strong assumptions that are definitely not true in general. The approximation can however be useful in building some intuition.

Consider the case where the portfolio only consists of one payer swap. Then the value of the risky swap becomes

$$\begin{aligned}\tilde{V}_P^B(t) &= V_P^B(t) \\ &\quad + \text{LGD}^B\mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau)[-V_P^B(\tau)]^+\mathbb{1}_{\{t<\tau\leq T\}}\mathbb{1}_{\{\tau=\tau^B\}}\right] \\ &\quad - \text{LGD}^C\mathbb{E}_t^{\mathbf{Q}}\left[D(t,\tau)[V_P^B(\tau)]^+\mathbb{1}_{\{t<\tau\leq T\}}\mathbb{1}_{\{\tau=\tau^C\}}\right]\end{aligned} \tag{4.41}$$

Using the approximation of independence and postponed default equation 4.39 simplifies to

$$\text{CVA}^B(t) \approx \text{LGD}^B \sum_{j=1}^{n} V_{P+}^B(T_j)\mathbb{P}_t^{\mathbf{Q}}\left[T_{j-1} < \tau \leq T_j, \tau = \tau^C\right] \tag{4.42}$$

We see that CVA is a portfolio of payer swaptions weighted by the probability of them being 'exercised' by default of the counterparty. Equation 4.40 can be reduced to

$$\text{DVA}^B(t) \approx \text{LGD}^B \sum_{j=1}^{n} V_{R+}^B(T_j)\mathbb{P}_t^{\mathbf{Q}}\left[T_{j-1} < \tau \leq T_j, \tau = \tau^B\right] \tag{4.43}$$

We see that DVA has a similar interpretation as CVA: In this case a portfolio of receiver swaptions. Now the expression only depend on the default probabilities and swaption prices. Default probabilities can be stripped from CDS spreads and swaption prices can be observed in the market, so in this case the approximation simplifies the problem significantly.

## 4.4 Mitigating Credit Risk

The netting set mentioned earlier is a method for reducing credit risk. Other non-trivial cases are listed below.

### 4.4.1 Collateral

For OTC derivatives the typical way to mitigate credit risk is to agree to post collateral. The collateral rules are defined in the *Credit Support Annex* (CSA) to the ISDA Master Agreement.

Define the collateral posted by the counterparty at time $t$ as $C(t)$. A negative amount means that the bank has posted collateral to the counterparty. The EAD can then be written as

$$\text{EAD} = (\Pi(\tau) - C(\tau))^{+} \tag{4.44}$$

If $C(t) = \Pi(t)$ then there is clearly no credit risk, but this is quite unrealistic. Collateral is posted periodically, and even if it was posted continuously there would be a *cure period*, which is the time before default where the counterparty fails to post collateral. This period is typically assumed to be between 10 and 25 days [Hull and White (2012)]. Denote the cure period as $\Delta\tau$. Then the collateral posted would be $C(\tau - \Delta\tau)$, i.e. based on the market value at time $\tau - \Delta\tau$.

It is also uneconomical to transfer small amounts back and forth, so *minimum transfer amounts* are defined in the CSA, which is another reason for $C(t) \neq \Pi(t)$.

Another possible feature of the CSA is the *threshold* that we denote by $\underline{C}$, which is an amount of credit risk that the parties are willing to bear. In this case the collateral function becomes

$$C(t) = \begin{cases} \Pi(t) - \underline{C} & \text{if } \Pi(t) > 0 \\ \Pi(t) + \underline{C} & \text{if } \Pi(t) < 0 \end{cases} \tag{4.45}$$

A negative threshold is called an *independent amount*.

It is clear that collateral does not remove credit risk altogether, but it also creates some other issues. If collateral re-hypothecation is allowed then the collateral taker have unrestricted use of the collateral until it must be returned to the collateral provider. There is thus a probability that other creditors will have a claim towards the collateral in case of default and the collateral provider can thus not expect to get the full amount back [Brigo, Morini and Pallavicini (2013), p. 309]. Additionally posting collateral is a cause for liquidity risk. It might be expensive to obtain cash or assets to post as collateral.

### 4.4.2 Central Counterparties

Central clearing of derivatives is another approach to mitigating credit risk. When dealing with central counterparties there is variation margin based on daily mark-to-market as well as initial margin. The variation margin plays the same role as the collateral in OTC transactions and the initial margin plays the same role as the independent amount [Brigo, Morini and Pallavicini (2013), p. 362]. So in that sense similar issues are present even though credit risk is mitigated.

### 4.4.3 Hedging

Standard CDSs are traded with a fixed notional. As the market value of the swap portfolio will fluctuate, a CDS with a notional that is contingent on some underlying index is needed. Contingent CDSs has historically been seen as a good hedge of CVA, but these also bear CVA as the dealer is not default risk-free and has limited liquidity, which reduces the effectiveness as a hedge [Brigo, Morini and Pallavicini (2013), p. 263].

Looking at the approximations in equation 4.42 and 4.43 gives the idea that we can hedge CVA and DVA on a single swap by entering a portfolio of swaptions. The portfolio weights are determined from the default probabilities. This holds when the portfolio consists only of a single swap and given independence. Using the approach described in section 3.2.2 and 3.4.2 it would be possible to determine the Greeks for CVA and DVA with respect to model parameters and market instruments. But as CVA and DVA is found using Monte Carlo simulation the 'bump and revalue'-approach is not feasible as every bump will start a new full Monte Carlo simulation making it extremely expensive to calculate the sensitivities.

A solution to this problem is to calculate the sensitivities simultaneously with the value itself using *Adjoint Automatic Differentiation* (AAD). The approach is easy in theory and is based on application of the chain rule, but difficult to implement in practice. There exists a version of `QuantLib`, where AAD is implemented called QuantLibAdjoint, but time did not permit trying to get this to work. An example of how the delta vector for a swap will look is given in this blog post by Peter Caspers.

In terms of sensitivities to the market instruments CVA and DVA depend on swap rates, swaption volatilities, CDS spreads and CDSO volatilities. So continuous trading in these instruments should in theory make it possible to hedge CVA and DVA, but without a model for the sensitivities this is practically impossible.

# 5

# Numerical Examples

This chapter will showcase the practical implementation of the theory. The code has been implemented in `C++` using the open source financial library `QuantLib`. Technical aspects are in appendix A.1. The source code for my implementation is available in my Dropbox.

## 5.1 Implementation Design

In the implementation we approximate the realised ZCB over the short time interval $\Delta t$ as

$$e^{-\int_t^{t+\Delta t} r(u)\,\mathrm{d}u} \approx \mathbb{E}_t^{\mathbf{Q}}\left[e^{-\int_t^{t+\Delta t} r(u)\,\mathrm{d}u}\right] = P(t, t+\Delta t)$$

As all the models used have a function for the discount bond this is very easy to implement. This is not exactly correct, but should be fairly close for small $\Delta t$. With $n$ discretisation points between $t_0$ and $t_n$ the numeraire of the risk-neutral measure is approximated using

$$\beta(t_n) \approx \beta(t_0)\frac{1}{\prod_{i=1}^{n} P(t_{i-1}, t_i)}$$

where $\beta(t_0) = 1$. An equivalent approximation method is used for the intensity with $r(t)$ replaced by $\lambda(t)$. To find the default time, we need to find the first $t$, where the integrated intensity is higher than an independent exponential variable. As mentioned above the approximation is used

$$e^{-\int_t^{t+\Delta t} \lambda(s)\,\mathrm{d}s} \approx \prod_{i=1}^{n} P^{CIR}(t_{i-1}, t_i)$$

So with $t$ being one of the $n+1$ discretisation points we need to find the first time at which below condition is true

$$\xi \leq \int_{t_0}^{t} \lambda(s)\,\mathrm{d}s \qquad \Longleftrightarrow \qquad e^{-\xi} \geq \prod_{i=1}^{n} P^{CIR}(t_{i-1}, t_i)$$

To find the standard exponential random variable we will use the inverse transform method described in section 2.3. The inverse cumulative distribution function is given by

$$F_{\xi}^{-1}(u) = -\log(1-u)$$

where $u$ is obviously between 0 and 1. We can thus simplify the expression further

$$\xi \leq \int_{t_0}^{t} \lambda(s)\,\mathrm{d}s \qquad \Longleftrightarrow \qquad 1 - U \geq \prod_{i=1}^{n} P^{CIR}(t_{i-1}, t_i)$$

where $U$ is uniform on $[0,1]$ and independent of all the other variables. This leads us to the result

$$\tau = \inf\left\{t : 1 - U \geq \prod_{i=1}^{n} P^{CIR}(t_{i-1}, t_i)\right\} \tag{5.1}$$

This expression is particularly useful as we can find the survival probability directly from the object. For a path we will thus have the following approach

1. Draw one uniform $U$
2. On every step of the path
   (a) Simulate correlated standard normals using Cholesky decomposition.
   (b) Update the bank account and survival probabilities using the approximation.
   (c) Check condition in equation 5.1. If true set $\tau = t_i$ and break loop. If false continue loop.

We can now have two possible outcomes: If default happened ($\tau \leq T$) we calculate the loss at $\tau$ and discount it back to time $t$. If default did not happen the CVA for the path is zero. DVA is calculated simultaneously and that should not have a major impact on efficiency.

It should be noted that there are some typos in the program, but unfortunately I did not have time to correct this. The bank account is set to $e^{-\int_{t_0}^{t_n} r(u)\,\mathrm{d}u}$ instead of $e^{\int_{t_0}^{t_n} r(u)\,\mathrm{d}u}$ meaning that when the payoff is divided by the bank account for discounting it it really multiplied by the bank account. Additional $\tau$ is set as $t$ instead of $t + \Delta t$, so this is also slightly inconsistent with the algorithm described above. For some paths it seems like the instantaneous forward rate 'explodes' meaning that the it becomes extremely negative (up to -30%). It has not been possible to find out why that is happening, but it is suspected that it is due to the linear interpolation not producing a smooth forward curve. The most recent Euribor fixing is approximated as the simple forward at the time of default.

## 5.2 Calibration to Recent Data

### 5.2.1 Interest Rate Calibration

Interest rate data is based on 3m Euribor. Rates are in percent and volatilities are in basis points. Data for 6m Euribor is very similar.

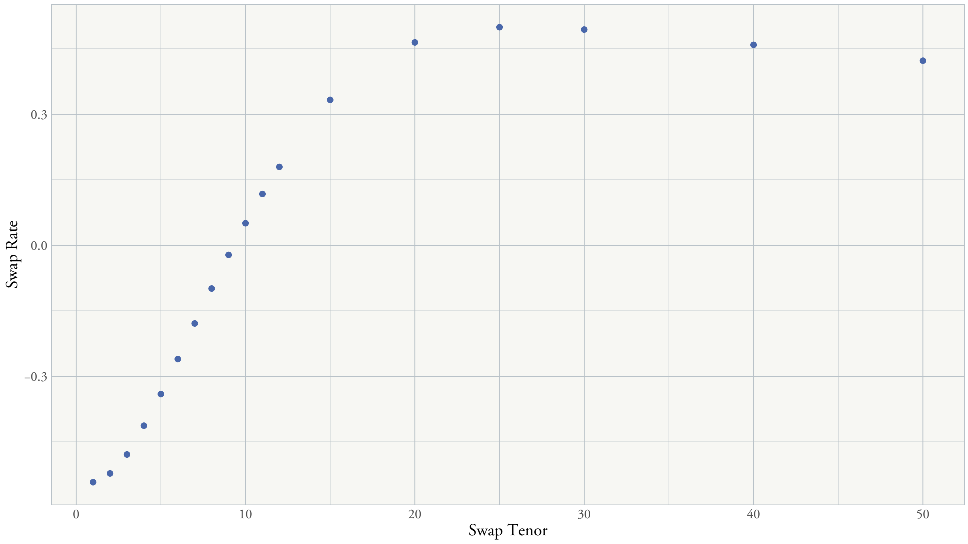



Figure 5.1: Swap Rates on 4th May 2021. Source: Bloomberg

This data will be used in constructing the zero curve. The bootstrapped zero curve is plotted below

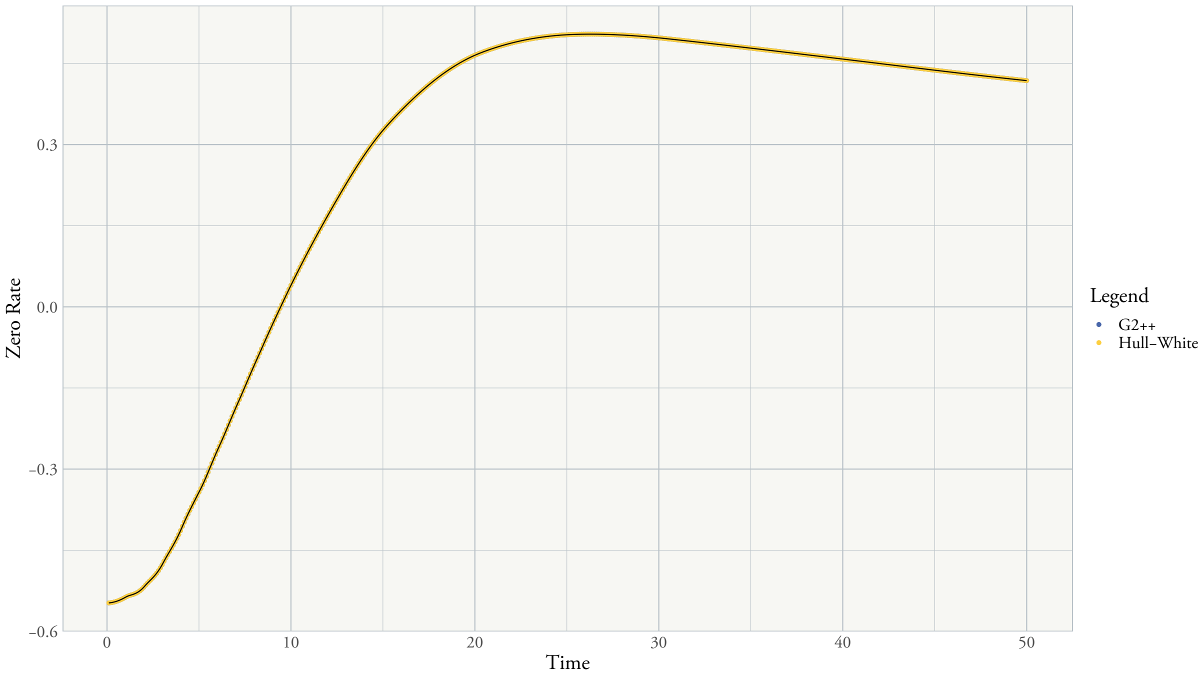



Figure 5.2: Bootstrapped Zero Curve and Model Implied Zero Rates

Both Hull-White and G2++ match exactly the bootstrapped zero curve.

The swaption volatilities are input with two decimals, but are rounded to nearest basis point in the below illustration

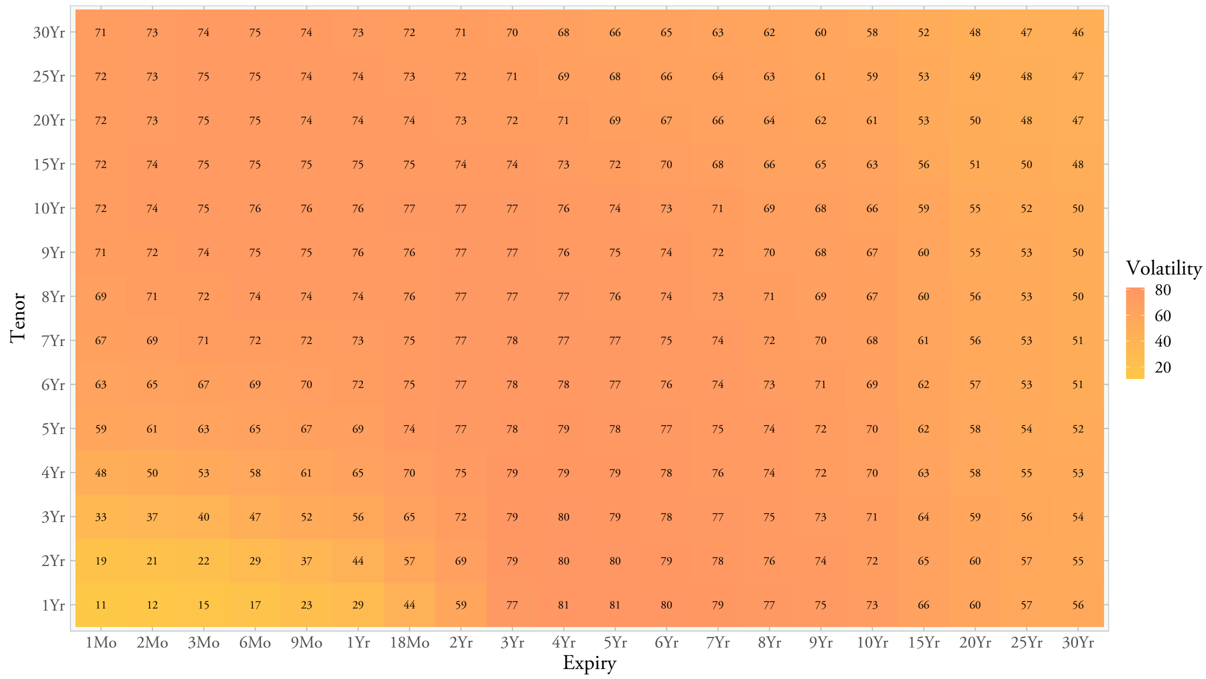



Figure 5.3: Swaption Normal Volatility Surface on $4^{th}$ May 2021. Source: Bloomberg

Implied volatility are highest for swaptions with maturity around 8 years from now and are low for very short-dated and very long-dated swaptions.

Some strange issues with the 1m/8y and 1m/9y swaptions have been experienced when trying to calibrate the G2++ model. Technically the model throws an error only for those two swaptions – everything else works fine. The error has been investigated unsuccessfully, so the solution is to disregard the two swaptions in this calibration. Using Swaption prices we try to calibrate two different models: Hull-White and G2++.

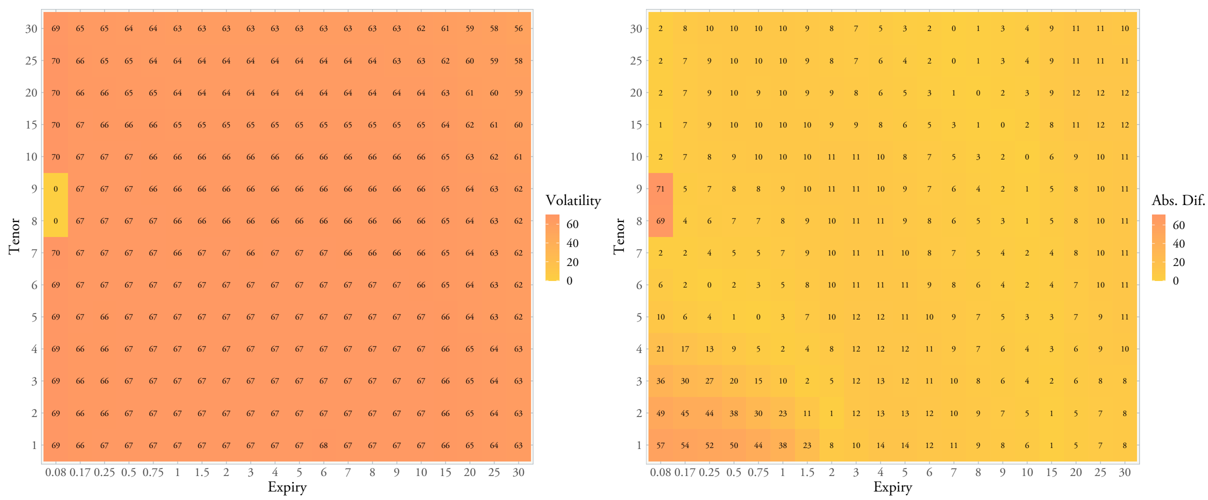



Figure 5.4: Calibration in G2++

In the Hull-White calibration all swaptions work, so here nothing is disregarded.

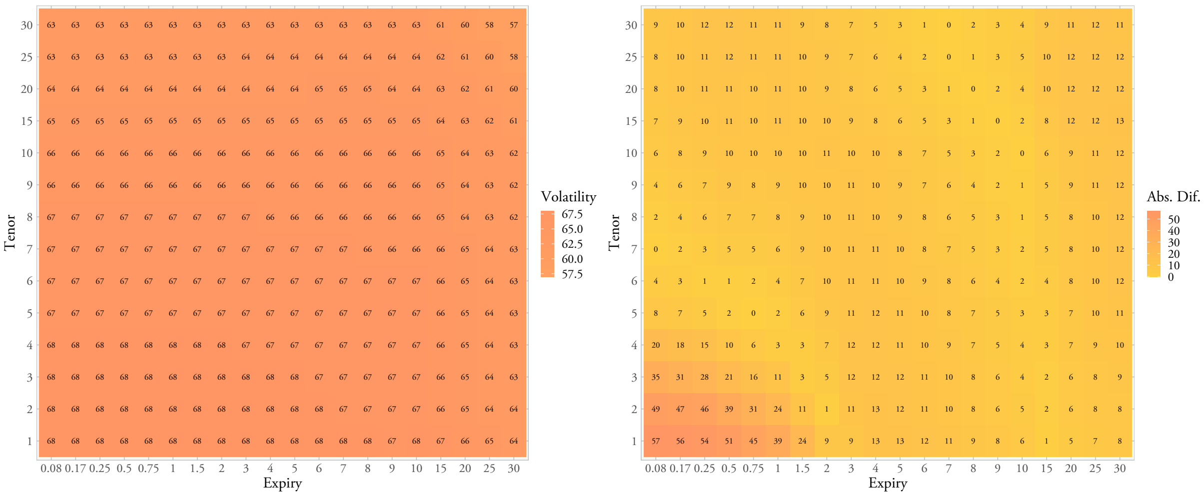



Figure 5.5: Calibration in Hull-White

The sum of absolute differences is 2 724 for the G2++ and 2 677 for the Hull-White model, so the fit is similar for the two models. If the 1m/8y and 1m/9y differences are ignored the values become 2 584 and 2 671, respectively, so the G2++ is deemed to fit the swaption data marginally better.

Parameters in the calibrated G2++ model is given by

$$\kappa_1 = 0.00549236, \quad \sigma_1 = 0.00970193, \quad \kappa_2 = 0.00550213, \quad \sigma_2 = 0.00955238, \quad \rho_{1,2} = -0.759051$$

and for the calibrated Hull-White model the parameters are

$$\kappa = 0.00531006, \quad \sigma = 0.00673971$$

runtimes are listed below

```
G2: Calibration run time = 1.07207e+06 ms
Hull-White: Calibration run time = 25986.3 ms
```

It is seen that the G2++ takes around 16 minutes to calibrate, while the Hull-White model only takes 26 seconds. The G2++ model is thus 41 times slower than the Hull-White model.

The implied volatility of the model prices is pretty homogenous and it seems like the models have difficulty replicating the volatility skew seen in the market – especially for swaptions close to expiry. This is most likely due to the short rate being Gaussian and having time-homogenous volatility due to constancy of the mean reversion and diffusion parameters.

The benefit the marginally better fit to data in G++ comes with a strange error for some swaptions and significantly larger runtime. This makes Hull-White a more robust choice of model, but the calibrated G2++ will anyway be used going forward due to larger consistency with market data.

### 5.2.2 Intensity Calibration

Ideally in calibration of the default probability curve we would have liquid debt instruments as a bond or CDS on the counterparty. As CDSs follows standards and the market is more transparent and liquid it is typically preferable to calibrate to CDSs than bonds. There exists a basis between bonds and CDSs due to the same reason. But in the case where the counterparty does not have liquid bonds or CDSs trading, a proxy for the credit risk associated with trading with the counterparty could be an average of companies in the same region and industry sector and with the same credit rating[17]. Unless the company is financially distressed, this will probably be a good proxy for the credit risk. IHS Markit provides this data under the name of CDS Sector Curves. The bank is modelled using the 'Financials' industry sector and the counterparty is modelled using the 'Corporates' industry sector to make sure that CDS's on governments is not included in the average.

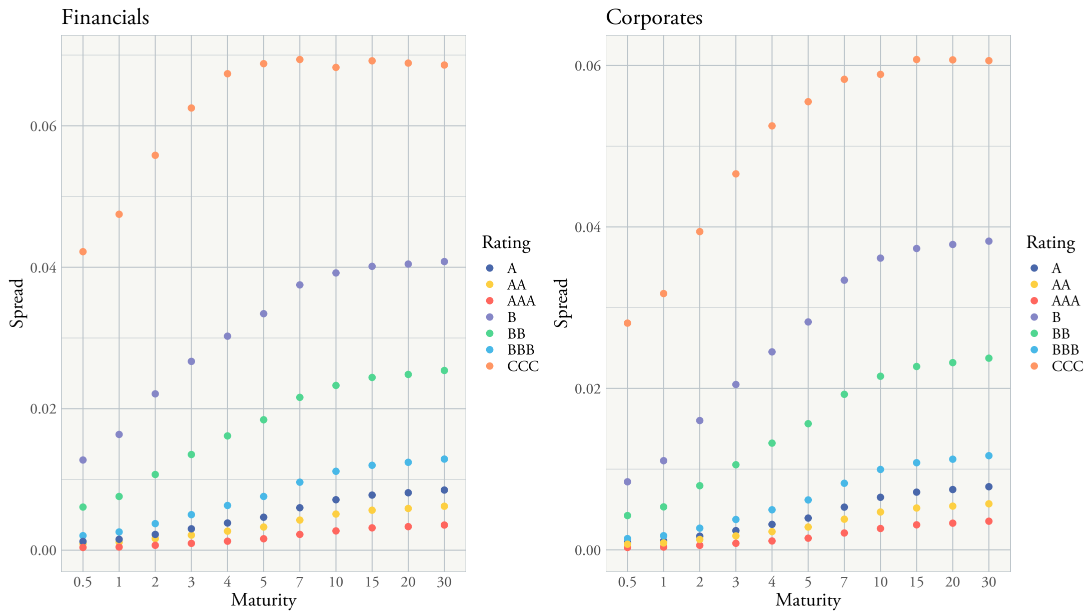



Figure 5.6: CDS Sector Curves on 4th May 2021. Source: IHS Markit

It is seen that the credit spread for financials is a bit higher than the corporates of similar credit quality. To be concrete it will be assumed that the bank has a credit rating of 'A' and that the counterparty has a credit rating of 'BBB'. This means that the counterparty has a slightly worse credit quality than the bank has. By using the above curve it is assumed that the companies can be represented as the average of the group. The spreads are based on CDSs with senior debt as underlying and under the assumption of 40% recovery. The bootstrapped survival probabilities is compared to the calibrated CIR++ model below

[17]The credit ratings follows the same scale as Standard & Poor's. Details on the scale and credit rating in general can be found here.

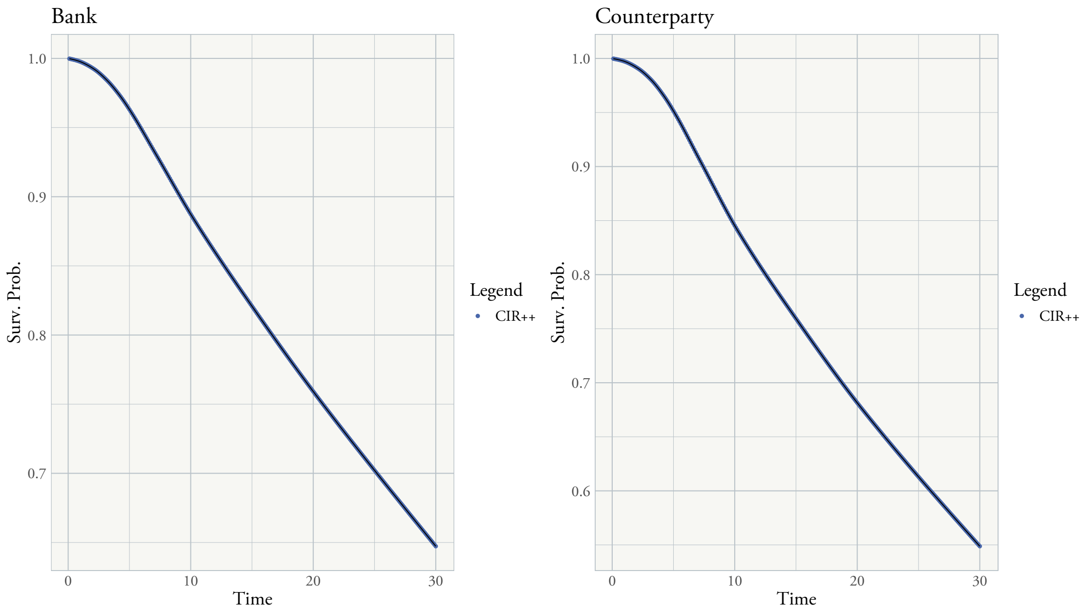



Figure 5.7: Survival Probabilities for the Bank and Counterparty

It has not been possible to obtain prices on CDS options. This is problematic as it is not possible to calibrate the intensity only to credit spreads. Of course one could try to estimate the parameters, but this would by definition be backward-looking, which is contrary to calibration, which is forward-looking by design. Using the CIR++ model it is possible to exactly match observed credit spreads, while setting the constant parameters arbitrarily. To be able to make a qualified guess on what the parameters should be, the realised volatility on the 'iTraxx Europe' Credit Default Swap Index with a maturity of 5 years has been investigated[18]. The contract comprises most liquid single name CDS's in Europe and can thus be seen as a benchmark to credit risk in Europe. It should be noted that the volatility of an average is lower than the average volatility of its constituents, so the volatility on single name CDS contracts will probably be larger than below.

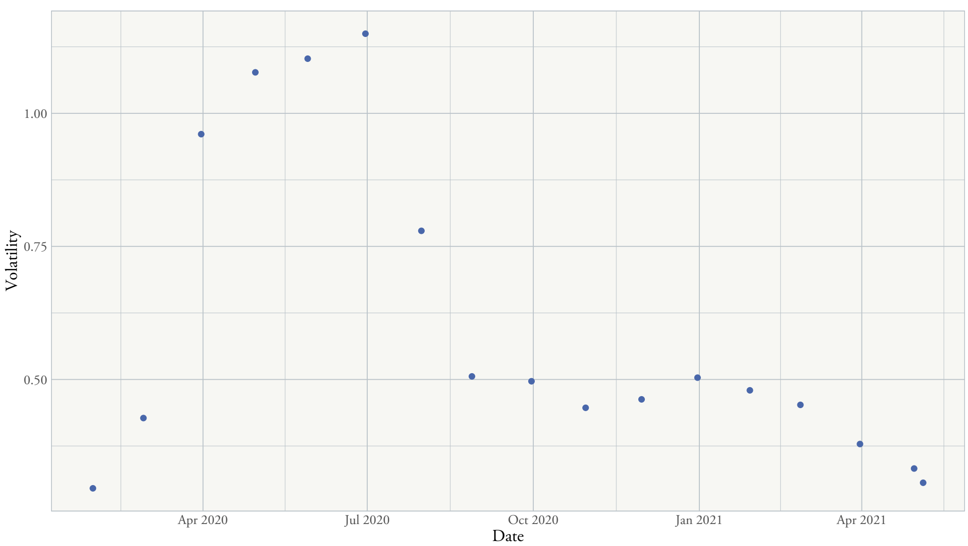



Figure 5.8: Realised Volatility on 5y Index CDS. Source: IHS Markit

[18]The 5 year contract is the most liquid of the contract, hence the choice of this contract. See link for more information.

The realised volatility is an annualised 90-day rolling average. It is seen that the Covid-19 pandemic increased credit spread volatility drastically. The credit spread volatility is now relatively low and on approximately the same level as before the pandemic. Usually realised volatility is below implied volatility due to a risk-premium being paid from option buyer to option seller. If you only charged the 'fair' volatility based on realised volatility, you are actually not accounting for black swan events.

The parameters from [Brigo, Morini and Pallavicini (2013), p. 125] will be used:

$$y(0) = 0.0165, \quad \kappa = 0.4, \quad \theta = 0.026, \quad \sigma = 0.14$$

And the dynamics of the model is given by

$$\lambda(t) = y(t) + \varphi(t) \tag{5.2}$$

where

$$\mathrm{d}y(t) = \kappa(\theta - y(0))\,\mathrm{d}t + \sigma\sqrt{y(t)}\,\mathrm{d}W(t) \tag{5.3}$$

From equation 3.124 we know the variance of the CIR process. In this concrete case the yearly standard deviation is 1.57%, which is 95.42% of $y(0)$, so this seems reasonable based on the illustration of realised volatility. It should be noted that volatility in hazard rates is different from volatility in CDS spreads, so it is not entirely comparable, but the two quantities should be highly correlated. It could have been interesting to look at the implied volatilities on CDS Options that is implied by this model, but the Monte Carlo simulation produced a very strange volatility surface[19], so it has been decided not to include this. Together with the issues mentioned earlier, this makes it pretty clear that there is something wrong with the implementation.

## 5.3 Credit Risk Calculations

For simplicity only portfolios of a single spot-starting payer swap will be considered, but the implemented function will also work for a portfolio of swaps. Maturities are set to span 5 years to 30 years with 5 year increments. The fixed rate is set as ATM±25bp. The case where the bank is *in the money* is denoted ITM and the case where the bank is *out of the money* is denoted OTM.

Three scenarios are considered: Independence, WWR and RWR. It could have been interesting to compare the independence scenario with the closed-form approximation described in section 4.3.4. The formulas are implemented in `QuantLib`, but I was not able to get it so produce sensible numbers for the example below[20].

As the counterparty is receiving fixed in the swap, rising interest rates would have a negative impact on the counterparty. WWR would then be when increasing interest rates coincides with increasing intensities, which can be obtained by assuming positive correlation between the two. The extreme case of correlation equal to 1 is chosen to illustrate WWR, while RWR is illustrated by setting correlation equal to -1. The swaps used in the example is listed below

[19]The 6m/6m CDSO had an implied volatility of 98%, while the 3y/6m had 15%.

[20]CVA was zero and BVA was thus very positive, so clearly something was wrong with my code.

| Maturity | ATM | | ITM | | OTM | |
|---|---|---|---|---|---|---|
| | Rate | NPV | Rate | NPV | Rate | NPV |
| 5 | -34 | 0 | -59 | 12 662 | -9 | -12 662 |
| 10 | 5 | 0 | -20 | 25 246 | 30 | -25 246 |
| 15 | 33 | 0 | 8 | 37 357 | 58 | -37 357 |
| 20 | 46 | 0 | 21 | 48 926 | 71 | -48 926 |
| 25 | 50 | 0 | 25 | 60 078 | 75 | -60 078 |
| 30 | 49 | 0 | 24 | 70 933 | 74 | -70 933 |

Table 5.1: Default Risk-Free Valuation

The time increment, $\Delta t$, is set to $\frac{1}{52}$ meaning that default is checked once per week. Increasing the number of time steps did not hurt performance significantly – even when compared to only having one time step per year this was only twice as slow for a 5 year swap.

### 5.3.1 G2++

The parameters of the calibrated G2++ model shown in section 5.2.1 will be used. When the interest rate is modelled by a G2++ process then the correlation between the interest rate and intensity dynamics is given by

$$\mathrm{Corr}(\mathrm{d}r(t), \mathrm{d}\lambda(t)) = \frac{\sigma_1 \rho_{1,3} + \sigma_2 \rho_{2,3}}{\sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho_{1,2}}} \tag{5.4}$$

This is shown in [Brigo, Morini and Pallavicini (2013), p. 126]. To obtain the wanted correlation we have to choose appropriate levels for $\rho_{1,3}$ and $\rho_{2,3}$. For the case of WWR $\rho_{1,3}$ is set (arbitrarily) to -0.05 and $\rho_{2,3}$ is set to 0.7505596 such that $\mathrm{Corr}(\mathrm{d}r(t), \mathrm{d}\lambda(t)) = 1$. For the case of RWR $\rho_{1,3}$ is again set to -0.05 and for $\mathrm{Corr}(\mathrm{d}r(t), \mathrm{d}\lambda(t))$ to equal, $-1$ $\rho_{2,3}$ is set to -0.648994. For simplicity it is assumed that the bank's credit intensity is uncorrelated with the interest rate. We also assume that the intensity process of the bank is uncorrelated with the counterparty's intensity process for simplicity.

The value in parenthesis is $\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \frac{s_n}{\sqrt{n}}$ as defined in section 2.3, where $\alpha$ is set to 0.95. The value plus/minus the parenthesis is thus the confidence interval.

It takes around 100 minutes to calculate CVA and DVA for 18 swaps using 10 000 paths. Concrete runtimes are listed below

```
Independent: Run time = 4.98817e+06 ms
Wrong Way Risk: Run time = 5.28099e+06 ms
Right Way Risk: Run time = 5.46163e+06 ms
```

This means that it takes around 6 minutes to run 10 000 simulations, which is by no means impressive. If time had permitted optimisation of runtime would have been investigated.

| Maturity | ATM | | ITM | | OTM | |
|---|---|---|---|---|---|---|
| | CVA | DVA | CVA | DVA | CVA | DVA |
| 5 | 478 (48) | 685 (171) | 683 (60) | 351 (44) | 311 (44) | 876 (77) |
| 10 | 2 043 (155) | 3 275 (232) | 2 702 (183) | 2 525 (217) | 1 345 (124) | 4 342 (258) |
| 15 | 5 120 (325) | 8 067 (459) | 7 249 (404) | 5 936 (410) | 3 779 (272) | 9 637 (520) |
| 20 | 11 052 (593) | 11 986 (641) | 13 597 (662) | 9 279 (551) | 8 163 (495) | 14 973 (722) |
| 25 | 19 537 (936) | 16 435 (882) | 24 514 (1 046) | 12 225 (748) | 16 160 (826) | 19 398 (953) |
| 30 | 30 878 (1 334) | 19 860 (1 101) | 37 358 (1 449) | 15 213 (936) | 27 418 (1 271) | 24 496 (1 247) |

Table 5.2: Value Adjustments under Independence

For the case where interest rates and intensities are independent it is seen that for long maturities CVA is typically larger than DVA, which makes good sense as the counterparty is assumed to have worse credit quality than the bank. It is seen that when the swap is ITM CVA increases while DVA decreases compared to when the swap is ATM and vice versa for the OTM case. If the swap have positive value for the bank initially then the probability of the swap having a positive value at the time of default, so this is consistent with intuition. It is quite puzzling that the short maturity ATM swaps have higher DVA than CVA, but otherwise the table is consistent with intuition. In general the confidence intervals are quite large, but this is expected as only 10 000 paths were made per simulation.

| Maturity | ATM | | ITM | | OTM | |
|---|---|---|---|---|---|---|
| | CVA | DVA | CVA | DVA | CVA | DVA |
| 5 | 828 (67) | 555 (58) | 1 237 (90) | 393 (55) | 563 (59) | 894 (76) |
| 10 | 3 356 (211) | 3 572 (250) | 5 022 (273) | 2 667 (207) | 2 610 (184) | 4 395 (264) |
| 15 | 8 832 (444) | 8 191 (439) | 10 515 (500) | 6 350 (379) | 6 977 (392) | 9 930 (504) |
| 20 | 17 594 (760) | 13 351 (689) | 21 341 (861) | 9 975 (581) | 14 317 (676) | 16 017 (763) |
| 25 | 28 657 (1 123) | 17 089 (905) | 32 291 (1 201) | 13 635 (768) | 23 803 (1 011) | 22 120 (1 054) |
| 30 | 43 447 (1 553) | 21 004 (1 125) | 50 687 (1 715) | 16 845 (1 007) | 37 106 (1 444) | 27 228 (1 300) |

Table 5.3: Value Adjustments under Wrong Way Risk

For the case where interest rates and intensities are positively correlated (WWR) it is seen that DVA is pretty similar to the case of independence. As the intensity of the bank is assumed to be uncorrelated with interest rates this makes good sense. When taking the confidence intervals into account it seems like it is probable that true value of DVA is equal in the two cases for most of the swaps. The intuition tells that DVA should be slightly lower under WWR as default of the counterparty set DVA to zero for that path, so it is strange that the values in general are higher with WWR than without.

CVA is much larger under WWR, which is logical. It is seen that WWR has a similar impact on ATM and ITM swaps and it has a larger impact on OTM swaps. Figure 2.12 can help in building some intuition for this. With positive correlation it is seen that cases where interest rates increase and intensities decrease become less likely meaning that for the scenarios, where the contract goes from OTM to ITM, have a higher probability of coinciding with financial distress of the counterparty. If rates increase then intensities must also increase due to the high positive correlation. This can also explain the larger DVA as DVA is highest when rates decrease and in the cases with decreasing rates the intensity of the counterparty decreases, making default of the counterparty less likely to occur when the bank is OTM thus increasing DVA.

| Maturity | ATM | | ITM | | OTM | |
|---|---|---|---|---|---|---|
| | CVA | DVA | CVA | DVA | CVA | DVA |
| 5 | 254 (42) | 726 (176) | 427 (55) | 375 (56) | 113 (20) | 973 (212) |
| 10 | 984 (104) | 3 284 (243) | 1 465 (126) | 2 318 (185) | 591 (73) | 4 290 (262) |
| 15 | 2 931 (232) | 7 016 (415) | 4 187 (289) | 5 830 (373) | 1 891 (182) | 9 376 (513) |
| 20 | 6 821 (450) | 11 058 (620) | 8 493 (504) | 8 578 (544) | 4 644 (362) | 14 045 (702) |
| 25 | 13 636 (772) | 14 641 (795) | 16 977 (855) | 11 086 (695) | 10 128 (643) | 18 063 (926) |
| 30 | 22 878 (1 162) | 17 218 (982) | 27 347 (1 254) | 13 961 (918) | 18 636 (1 032) | 22 555 (1 143) |

Table 5.4: Value Adjustments under Right Way Risk

Introducing RWR has the opposite impact on the value adjustment than WWR. Using the plot of negatively correlated variables in 2.12 we can again explain the reduction in DVA as a consequence of right way risk even though the bank's intensity is uncorrelated with interest rates.

It is seen from above numerical examples that the impact of WWR can be material sometimes doubling the BVA. Assuming zero correlation can thus lead to significant error in valuation, which can lead to credit losses that could have been avoided, if the bank had charged the counterparty for this risk at inception.

## 5.4 Case Study: Swaps in Housing Cooperatives

In Denmark interest rate swaps are somewhat infamous as media coverage typically involves some company being in financial distress due to interest rate swaps traded before the financial crisis, where rates were much higher than they are today. Typically the bank is criticised for having advised the company to enter a swap and recently there has been some critique of the way banks calculate the market value of the swaps.

An odd place for the use of interest rate swaps has been in housing cooperatives, where the swaps have been used to convert a variable interest rate into a fixed one. Without credit and spread risk the construction is perfectly reasonable. The variable rate received in the swap should cancel out the variable rate on the bond leaving the housing cooperative with only a fixed rate to pay. Looking at the interest rate on the so-called 'F1' mortgage bond, the interest rate seems a bit higher than Euribor. The F1 is a floating rate bond, where the rate is adjusted yearly and the current level is around -20bp. Compared to the 1 year swap rate of around -50bp for both 3m and 6m Euribor there is quite a difference between the two rates. If the spread between F1 and Euribor is constant this is not an issue per se as the 30bp could be seen as a risk-premium to be added on top of the fixed rate payed in the swap, but if the spread is not constant it is also a source of risk for the housing cooperative.

A particularly interesting example is the case of 'Andelsboligforeningen Hostrups Have', which was the largest housing cooperative in Denmark at the time of default in December 2016. Details are described in [Kailay (2017)]. The article is in Danish and behind a paywall, so the key facts are listed below. The housing cooperative was founded in 2007 with the purchase of Hostrups Have for 1.1 billion DKK. They financed the purchase through a variable rate mortgage, which was then swapped to a fixed rate. The fixed rate in the swap was set to be in 3.65% in 2007 rising to 6.2% in 2028. This structure gives associations to the infamous *teaser rates* of the sub-prime mortgage crisis. According to a former board member it was 'too expensive' to take on a fixed rate mortgage and 'too risky' to take on a variable rate mortgage only, which was the justification for choosing above construction.

The housing cooperative was declared bankrupt by the bank, Nykredit, in December 2016 after it decided to lower the rent making it improbable that the housing cooperative would be able to fulfil its obligations towards Nykredit. At the time of default the notional on the mortgage was 1.1 billion DKK suggesting that it was an interest-only mortgage. The swap had a negative market value of 900 million DKK and the property was only valued at only 745 million DKK. Nykredit lost 250 million DKK and the shareholders of the cooperative lost the equity they had originally invested in the purchase.

This example should make it pretty clear that CVA is material and can amount to huge losses for the bank. This loss can be seen as a DVA gain for the shareholders in the cooperative as they did not have to pay Nykredit back in full. To understand how it went so wrong for the housing cooperative the term structure and development of swap rates are illustrated

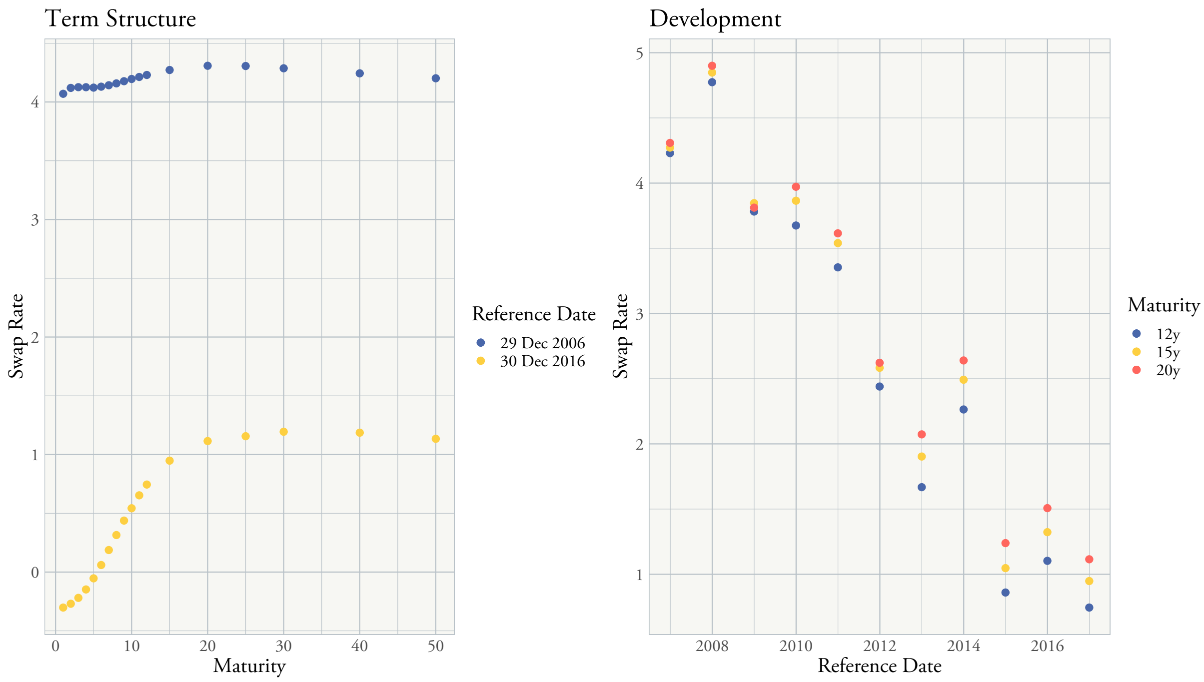



Figure 5.9: Swap Rates for 3m Euribor

The swap rates have decreased dramatically through the life of the swap. With a flat discounting curve of 4.5% the value of the fixed rate staircase is approximately the same value as paying 4.5% fixed suggesting that the deal was fair initially based on above illustration of the swap rate term structure. At the time of default the housing cooperative was paying a bit below 5% fixed, while the market 12y swap rate was 0.74%. The average rate of the last 12 years of the loan was around 5.5%, so with an annuity of approximately 12[21] this amount to a loss of $\approx (0.0074 - 0.055) \cdot 12 = -0.5712$ per unit notional if the swap had been entered with a default risk-free counterparty. As the notional of the swap was probably constant at 1.1 billion DKK this amounts to 628 million DKK. This is a very approximate calculation, but it is done to illustrate that the loss of 900 million DKK was probably not an exaggeration. Nykredit has a credit loss of 250 million DKK, so if credit risk is taken into account this does not seem too far off.

With hindsight it is easy to see that the housing cooperative should not have entered the swap. The reason for the fixed rate bond being more expensive than the construction chosen by the cooperative is due to the embedded option to always be able to pay back the loan at par. With falling interest rates it is thus possible to 'convert' the mortgage to a lower interest rate. This optionality would have had a huge value in the case. Alternatively a 5y15y payer swaption could have been entered to protect the housing cooperative from rising interest rates within the first 5 years after they were founded. The premium paid on the swaption would have been lost, but the cooperative would have avoided financial distress.

Nykredit have probably hedged their interest rate risk by entering a swap with another bank, where Nykredit pay a lower rate than the counterparty pay on their swap, so Nykredit is able to pocket the spread as long

[21] The annuity $A(t)$ is equal to the maturity when the zero rate is constant at zero and approximately when rates are low.

as the counterparty is fulfilling its obligations towards Nykredit. At the time of default the hedge turned into an outright position as the exposure had disappeared. So instead of being market neutral, Nykredit was positioned for higher rates in a time of decreasing interest rate leading to huge mark-to-market losses. Using the theory from this thesis it could have been partly avoided if Nykredit had delta-hedged its interest rate exposure stemming from CVA. Nykredit would however not have been able to hedge its credit risk directly as CDSs on the counterparty is not a tradable product. Proxy hedging would not have helped in this case as the counterparty actually defaulted (jump to default risk).

# 6

# Conclusion

In this thesis the modelling of counterparty credit risk for interest rate swaps has been investigated. First general arbitrage theory was explained, then theoretical prices for interest rate derivatives, credit derivatives and counterparty credit risk were derived and lastly Bilateral Credit Value Adjustment (BVA) was calculated using Monte Carlo simulation in three different cases of correlation structure between interest rates and default intensities; independence, Wrong Way Risk and Right Way Risk. BVA is in general large and Wrong Way Risk can have very negative consequences for swap dealers as is seen from the case study of interest rate swaps in housing cooperatives.

It is shown that when the parties in the swap can default, the value of the swap no longer depends only on the term structure of interest rates, but also depends on the joint dynamics of interest rates and default intensities. The interest rate is modelled by a two factor Gaussian short rate model denoted G2++, which is a special case of the Gaussian and Markov Heath-Jarrow-Morton model, and the default intensity is modelled by an extension to the Cox-Ingersoll-Ross process denoted CIR++. The interest rate model has been calibrated to swap and swaption prices and the default intensity model has been calibrated to CDS spreads with both models being able to match the market prices of linear products exactly, while reproducing the volatility surface only to some extent. The models used are thus consistent with recent market data, which is essential for a sound valuation. A closed-form approximation of BVA under the assumption of independence and postponed default show that BVA can be approximated by a portfolio of swaptions with portfolio weights being a function of the probability of default.

Wrong and Right Way Risk are introduced into the model by introducing correlation between the Wiener processes driving the interest rate and the Wiener process driving the default intensity of the counterparty. This approach seems to work as expected.

## 6.1 Further Research

It could be interesting to investigate whether extending the model by assuming that the interest rate follows a Quasi-Gaussian model instead has significant impact on the counterparty credit risk. The Quasi-Gaussian model would probably have a better fit to market data and would thus provide more realistic dynamics for the interest rates.

Furthermore the other 'XVA's could be interesting to extend the model to with calculation of sensitivities using Adjoint Automatic Differentiation along the lines of [Huge and Savine (2017)].

# Bibliography


Andersen, Leif B.G. (Jan. 2007). *Efficient Simulation of the Heston Stochastic Volatility Model*. Working Paper. Available at SSRN: https://ssrn.com/abstract=946405.

Andersen, Leif B.G. and Vladimir V. Piterbarg (2010). *Interest Rate Modelling*. 1st ed. Atlantic Financial Press.

Björk, Tomas (2009). *Arbitrage Theory in Continuous Time*. 3rd ed. Oxford University Press.

Brigo, Damiano and Aurélien Alfonsi (Feb. 2004). *Credit Default Swaps Calibration and Option Pricing with the SSRD Stochastic Intensity and Interest-Rate Model*. Working Paper. Available at: https://cermics.enpc.fr/~alfonsi/cirppcredit.pdf.

Brigo, Damiano, Massimo Morini and Andrea Pallavicini (2013). *Counterparty Credit Risk, Collateral and Funding*. 1st ed. John Wiley & Sons, Ltd.

Cox, John C., Jonathan E. Ingersoll and Stephen A. Ross (1985). 'A Theory of the Term Structure of Interest Rates'. In: *Econometrica* 53.2, pp. 385–407. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1911242.

Derman, Emanuel and Paul Wilmott (Jan. 2009). *The Financial Modelers' Manifesto*. Working Paper. Available at SSRN: https://ssrn.com/abstract=1324878.

European Central Bank (Dec. 2020). *€STR Annual Methodology Review*. Technical Document. Available at: https://www.ecb.europa.eu/stats/financial_markets_and_interest_rates/euro_short-term_rate/html/ecb.eamr2012.en.pdf.

Fleming, Michael J. and Asani Sarkar (Dec. 2014). *The Failure Resolution of Lehman Brothers*. FRBNY Economic Policy Review. Available at: https://www.newyorkfed.org/medialibrary/media/research/epr/2014/1412flem.pdf.

Huge, Brian Norsk and Antoine Savine (June 2017). *LSM Reloaded - Differentiate xVA on your iPad Mini*. Working Paper. Available at: https://ssrn.com/abstract=2966155.

Hull, John and Alan White (2012). *CVA and Wrong Way Risk*. Working Paper. Available at http://www-2.rotman.utoronto.ca/~hull/downloadablepublications/WrongWayRisk.pdf.

ISDA (May 2021). *SwapsInfo First Quarter of 2021 Review*. Technical Document. Available at: https://www.isda.org/a/PMzTE/SwapsInfo-First-Quarter-of-2021-Review-Full-Report.pdf.

Jamshidian, Farshid (1989). 'An Exact Bond Option Formula'. In: *The Journal of Finance* 44.1, pp. 205–209. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2328284.

Kailay, Stefan Singh (21st June 2017). 'Svenskere køber Danmarks største andelsboligforening'. In: *Berlingske*. URL: https://www.berlingske.dk/virksomheder/svenskere-koeber-danmarks-stoerste-andelsboligforening (visited on 16/05/2021).

Karatzas, Ioannis and Steven E. Shreve (1988). *Brownian Motion and Stochastic Calculus*. 1st ed. Springer-Verlag.

Sorensen, Eric H. and Thierry F. Bollier (1994). 'Pricing Swap Default Risk'. In: *Financial Analysts Journal* 50.3, pp. 23–33. DOI: 10.2469/faj.v50.n3.23.

# A

# Appendix

## A.1 QuantLib

I am using QuantLib version `1.22-dev` with Boost version `1.75.0`. I have forked lballabio/QuantLib to mmencke/QuantLib on 29th March 2021. In broad terms I am using the functionality provided in the QuantLib library, but I am changing some implementations. I have installed the library using this guide, but where I build the library from my GitHub repository every time I change something in the source code.

To get the library working in Xcode version 9.2 I had to define search paths and linker flags

| | |
|---|---|
| Header Search Paths | `/usr/local/include` |
| Library Search Paths | `/usr/local/lib` |
| Other Linker Flags | `-lQuantLib` |

I have implemented calibration to CDS Options with inspiration from this link.

## A.2 Proofs

*Proof of Proposition 3.6.* Consider $X(t) = \ln S(t)$ with dynamics

$$\mathrm{d}X(t) = \frac{1}{S(t)}\,\mathrm{d}S(t) + \frac{1}{2}\left(-\frac{1}{S^2(t)}\right)(\mathrm{d}S(t))^2 = \sigma\,\mathrm{d}W(t) - \frac{1}{2}\sigma^2\,\mathrm{d}t \tag{A.1}$$

We can find the conditional value of $X(T)$ given $X(t)$:

$$X(T)|X(t) = X(t) + \int_t^T \sigma\,\mathrm{d}W(s) - \int_t^T \frac{1}{2}\sigma^2\,\mathrm{d}s = \sigma[W(T) - W(t)] + X(t) - \frac{1}{2}\sigma^2(T-t) \tag{A.2}$$

which is clearly normally distributed with mean $X(t) - \frac{1}{2}\sigma^2(T-t)$ and variance $\sigma^2(T-t)$. We can thus write the value of $S(T)$ given $S(t)$:

$$S(T)|S(t) \stackrel{d.}{=} S(t)e^{-\frac{1}{2}\sigma^2(T-t)+\sigma\sqrt{T-t}\varepsilon} \tag{A.3}$$

where $\varepsilon$ is a standard normal random variable. We can re-write the expected payoff

$$\mathbb{E}_t^{\mathbb{A}}\left[(S(T)-K)^+\right] = \mathbb{E}_t^{\mathbb{A}}\left[(S(T)-K)\,\mathbb{1}_{\{S(T)>K\}}\right] \tag{A.4}$$

And we see that under the annuity measure

$$S(T) > K \quad \iff \quad \sigma\sqrt{T-t}\varepsilon > \ln\frac{K}{S(t)} + \frac{1}{2}\sigma^2(T-t) \quad \iff \quad -\varepsilon < \frac{\ln\frac{S(t)}{K} - \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}} \tag{A.5}$$

Define $d_\pm = \frac{\ln\frac{S(t)}{K} \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}$. As the normal distribution is symmetric this is equivalent to

$$\varepsilon < d_- \tag{A.6}$$

So we can write

$$\mathbb{E}^{\mathbf{A}}\left[K\mathbb{1}_{\{S(T)>K\}}\right] = K \cdot \mathbb{P}^{\mathbf{A}}\left[S(T) > K\right] = K \cdot \mathbb{P}^{\mathbf{A}}\left[\varepsilon < d_-\right] = K \cdot \Phi(d_-) \tag{A.7}$$

We now consider the other part of the expectation, where we define $a := \sigma\sqrt{T-t}$

$$\begin{aligned}\mathbb{E}_t^{\mathbf{A}}\left[S(T)\mathbb{1}_{\{S(T)>K\}}\right] &= \mathbb{E}_t^{\mathbf{A}}\left[S(t)e^{-\frac{1}{2}a^2+a\varepsilon}\mathbb{1}_{\{\varepsilon<d_-\}}\right] \\ &= S(t)e^{-\frac{1}{2}a^2}\int_{-\infty}^{d_-} e^{ax}\phi(x)\,\mathrm{d}x \\ &= S(t)e^{-\frac{1}{2}a^2}\int_{-\infty}^{d_-} e^{ax}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\,\mathrm{d}x\end{aligned} \tag{A.8}$$

Consider now a change of variables to $z = x - a$ such that $x = z + a$. We can then write

$$\mathbb{E}_t^{\mathbf{A}}\left[S(T)\mathbb{1}_{\{S(T)>K\}}\right] = S(t)e^{-\frac{1}{2}a^2}\int_{-\infty}^{d_-+a} e^{a(z+a)}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z+a)^2}\,\mathrm{d}z \tag{A.9}$$

Wen can simplify the expression using $e^{-\frac{1}{2}a^2}e^{a(z+a)}e^{-\frac{1}{2}(z+a)^2} = e^{-\frac{1}{2}a^2}e^{az+a^2}e^{-\frac{1}{2}(z^2+a^2+2az)} = e^{-\frac{1}{2}z^2}$. Notice that $d_- + a = d_+$, so we can write

$$\begin{aligned}\mathbb{E}_t^{\mathbf{A}}\left[S(T)\mathbb{1}_{\{S(T)>K\}}\right] &= S(t)\int_{-\infty}^{d_+}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\,\mathrm{d}z \\ &= S(t)\Phi(d_+)\end{aligned} \tag{A.10}$$

It is now clear that the price of a swaption must be

$$V_{P+}(t) = A(t)\left[S(t)\Phi(d_+) - K\Phi(d_-)\right] \tag{A.11}$$

■

*Proof of Proposition 3.7.* This proof follows the same lines as that for the Black formula. We have the conditional value

$$S(T)|S(t) = S(t) + \int_t^T \sigma \, \mathrm{d}W(s) = S(t) + \sigma[W(T) - W(t)] \overset{dist.}{=} S(t) + \sigma\sqrt{T-t}\varepsilon \tag{A.12}$$

Again we have the payoff

$$\mathbb{E}_t^{\mathbf{A}}\left[(S(T)-K)^+\right] = \mathbb{E}_t^{\mathbf{A}}\left[(S(T)-K)\,\mathbb{1}_{\{S(T)>K\}}\right] \tag{A.13}$$

But now we have

$$S(T) > K \quad \Longleftrightarrow \quad S(t) + \sigma\sqrt{T-t}\varepsilon > K \quad \Longleftrightarrow \quad -\varepsilon < \frac{S(T)-K}{\sigma\sqrt{T-t}} \tag{A.14}$$

Define $d := \frac{S(T)-K}{\sigma\sqrt{T-t}}$. We note that $\phi'(x) = x\phi(x)$. We can then find the expected payoff with ease

$$\begin{aligned} \mathbb{E}_t^{\mathbf{A}}\left[(S(T)-K)^+\right] &= \int_{-\infty}^{d} \left[S(t) + \sigma\sqrt{T-t}x - K\right]\phi(x)\,\mathrm{d}x \\ &= [S(t)-K]\int_{-\infty}^{d} \phi(x)\,\mathrm{d}x + \sigma\sqrt{T-t}\int_{-\infty}^{d} \phi'(x)\,\mathrm{d}x \\ &= [S(t)-K]\Phi(d) + \sigma\sqrt{T-t}\phi(d) \end{aligned} \tag{A.15}$$

as $\lim_{x\to-\infty} \phi(x) = 0$. The swaption price must then be

$$V_{P+}(t) = A(t)\left[(S(t)-K)\,\Phi(d) + \sigma\sqrt{T-t}\phi(d)\right] \tag{A.16}$$

■