Dissertation

PATRÍCIA STARK

2022

Swaption Pricing Techniques in Different Interest Rate Models

PATRÍCIA STARK

2022

Economic and Financial Mathematical Analysis MSc
Financial Mathematics Specialisation
Corvinus University of Budapest
Supervisor: Prof. Péter Medvegyev

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Péter Medvegyev, for introducing me to probability theory and stochastic processes and guiding me with unique insights throughout my university years. He was the most influential teacher I have ever had.

A special thanks goes to Péter Paulovics, for proofreading, his helpful comments, and always being there for me when I needed advice.

And finally, I would like to thank my family, boyfriend and friends for loving and supporting me when I needed it most. I am grateful for the patience and kindness I received.

Contents

List of Tables

List of Figures

List of Algorithms

Chapter 1

Motivation

It is well known that the value of a payer swap at time t with start date T_0, payment dates T_i, for $i = 1, \ldots, n$ and a fixed coupon of c is

$$V_{swap}(t) = \sum_{i=0}^{n-1} \tau_i P(t, T_{i+1}) (L_i(t) - c)$$

where $\tau_i = T_{i+1} - T_i$ is the day count fraction between two consecutive payments, $L_i(t)$ is the floating leg, that is, the LIBOR rate, and $P(t, T)$ is the time t value of a T maturity zero coupon bond. For simplicity, we assume that the payment schedule of the two legs is the same, there is no spread, both legs have unit notional, and the fixed coupon is flat.

Suppose we have an option to enter into this swap at time T_0. This derivative is called a payer swaption, and its time t value is

$$V_{swaption}(t) = \beta(t) \mathbb{E}_t \left[\left(\sum_{i=0}^{n-1} \tau_i P(T_0, T_{i+1}) (L_i(T_0) - c) \right)^+ \beta(T_0)^{-1} \right]$$

where β is the numéraire, usually the money market account, and $\mathbb{E}_t$ is the conditional expectation taken under the risk-neutral measure corresponding to the numéraire.

It is obvious that the swaption's value function is the positive part of a sum, independent of our model choice. Being one of the most liquid products, swaptions play a crucial role in interest rate markets. They are particularly popular as a means of hedge. Consequently, it is paramount that their pricing be accurate and computationally efficient. Unfortunately, giving an explicit formula for this expectation is highly complicated, at least in its current form. It would be very convenient if we were to break this problem into smaller problems

– without making any simplifying assumptions – which are much easier to solve either analytically or numerically, depending on the model.

The Jamshidian (1989) decomposition does exactly that. Under certain general conditions, it decomposes the option of a sum into a basket – that is, a sum – of options. Needless to say, it is much more accurate if, for instance, a Monte Carlo pricer does not have to keep track of the many possible paths of the sum, only its components' (which often correlate or can even be priced analytically). It saves valuable computational power, time, and, by improving accuracy, money for the bank.

In this dissertation, we are interested in developing the Jamshidian decomposition and comparing it across different interest rate models. However, it only works if there exists an analytical formula for bonds (and, more precisely, for bond options as well). Therefore, our model choices are quite limited. There are only three model types where such formulae exist for both bonds and options: the Vasicek, the Hull–White, and the CIR. We are going to explore the decomposition in these models and compare it with two common numerical schemes for swaption pricing: Monte Carlo and tree methods.

The rest of the dissertation is organised as follows: in Chapter 2, we give an introduction to the basics of interest rates and the most important definitions. Chapter 3 is dedicated to swaps and swaptions. Chapter 4 presents the three swaption pricing techniques. In Chapter 5 we explore the Vasicek, Hull–White and CIR interest rate models. Having laid the theoretical foundations, in Chapter 6 we present the implementation of a Jamshidian, Monte Carlo and tree pricer for vanilla, accreter and amortiser swaptions. Our benchmark will be the Jamshidian engine, and we will examine the performance and parameter sensitivity of the other two methods. Finally, Chapter 7 concludes the dissertation.

Chapter 2

Introduction to Interest Rates

In this section, we establish the fundamental concepts of interest rate modelling and follow the notation and structure of Andersen & Piterbarg (2010).

Definition. A *zero coupon bond* is a security that pays a unit on its maturity T. Let $P(t,T)$ denote its price at time $t \leq T$.

We assume that $P(t,T)$ is $\mathscr{F}_s$-measurable, $\forall\ t \leq s \leq T$, where $\mathscr{F}_s$ is the filtration at time s. In other words, at any time $s \leq T$, we know what the zero coupon bond was worth from its start date to s. Obviously, from the definition it follows that $P(T,T)$ is $\mathscr{F}_s$-measurable $\forall s$, moreover $P(T,T) = 1$.

Notation. Let $P(t,T,T+\tau)$ denote the *forward price* of a zero coupon bond spanning from T to $T+\tau$ at time t, where $\tau > 0$.

Remark. In the absence of arbitrage, it is clear that

$$P(t,T,T+\tau) = \frac{P(t,T+\tau)}{P(t,T)}.$$

When talking about zero coupon bonds, deposits or even share prices, i.e. security prices, it is inevitable to introduce the concept of yields – or *rates* –, some kind of measure to capture how much the prices increased or fell during a period of time. It is important to note that a yield *always* corresponds to a time interval and cannot be interpreted itself. Many definitions of yields exist. Some are easy to grasp, others are more elusive, but no doubt all of them can play a useful part in explaining some properties of price movements. Here, we present the most commonly used rates in stochastic differential equations.

Definition. Let $y(t, T, T+\tau)$ denote the *continuously compounded forward yield* such that

$$e^{-y(t,T,T+\tau)\cdot\tau} = P(t, T, T+\tau).$$

Intuitively, the continuously compounded forward yield is a uniform exponential growth rate at which the price of the zero coupon bond increases over $[T, T+\tau]$ when it reaches 1.

It is important to note that if the price dynamics of the zero coupon bond is deterministic, then the continuously compounded forward yield is constant in its first variable t. However, most of the time this is not the case. Generally, zero coupon bond prices follow stochastic processes, so our *expectations* – which translate into yields or rates – about how the price will grow over $[T, T+\tau]$ are changing as we move forward in time, making t a true variable of $y(t, T, T+\tau)$.

Definition. In the previous definition, τ is called the *tenor* of the zero coupon bond, in other words, the length of the time interval between the start and maturity date.

Tenors are sometimes referred to as day count fractions when talking about interest rate products. There exist many conventions to calculate these, and they usually vary from currency to currency. The most popular products have annual, semiannual, quarterly, or monthly payment frequencies that put zero coupon bonds of these tenors under the spotlight.

Although continuously compounded rates are very straightforward and can be easily calculated theoretically, they are a little difficult to grasp for the general public. They also need exponential or logarithmic operations, which is computationally tiresome and can be approximated by linear functions reasonably well. That leads us to our next yield definition.

Definition. Let $L(t, T, T+\tau)$ denote the *simple forward rate* such that

$$P(t, T, T+\tau) = \frac{1}{1+\tau L(t, T, T+\tau)}.$$

Generally, the simple forward rate is assumed to be the LIBOR rate. We will keep this tradition throughout the work, although with the recent transition, things are getting a bit more complicated in reality.

Continuing the previous analogy, if the continuously compounded rate is a uniform, average-like measure of price movements (just like the average speed over a time interval), then we can define something like instantaneous speed at a point in time. The following definition aims to do that.

Definition. Let $f(t,T)$ denote the *instantaneous forward rate* spanning from t to T such that

$$f(t,T) = \lim_{\tau \searrow 0} L(t,T,T+\tau).$$

Remark. It is easy to show that

$$P(t,T,T+\tau) = \exp\left(-\int_T^{T+\tau} f(t,u)\,du\right).$$

Remark. From the previous remark, it is clear that

$$f(t,T) = -\frac{\partial \ln P(t,T)}{\partial T}.$$

Proof. We use $P(t,T) = P(t,t,T)$. The rest is simple and straightforward calculus. □

Remark. From the definitions of the continuously compounded and instantaneous forward rates, it is clear that

$$y(t,T,T+\tau) = \frac{1}{\tau}\int_T^{T+\tau} f(t,u)\,du \implies f(t,T) = \lim_{\tau \searrow 0} y(t,T,T+\tau).$$

The next important step is when not only the time span converges to zero, but the time to maturity too. Basically we described the overnight interest rate at time t:

Definition. Let $r(t)$ denote the *short rate* or *spot rate* at time t such that

$$r(t) = f(t,t).$$

Interest rate models choose one of these rates and model its dynamics. For instance, the Heath–Jarrow–Morton model works with the instantaneous forward rate, while numerous

models exist for short rates, like the Hull–White or Cox–Ingersoll–Ross models. The choice of rate is mainly a matter of convenience, and of course, by the previous remarks, one differential equation can be transformed into another.

As we work with swaps and swaptions, the following two definitions will also be useful, although they are not exactly rates.

Definition. An *annuity* is a contract that pays a unit at the end of each period between its start and end date T_k and T_{k+m}. The payment dates are denoted by T_i, and the accrual periods $[T_i, T_{i+1}]$ have length $\tau_i = T_{i+1} - T_i$, for $i = k+1, \ldots, k+m$.

Remark. If $A_{k,m}$ denotes the annuity starting at T_k with tenor m, then its fair price at time t is

$$A_{k,m}(t) = \sum_{i=k}^{k+m-1} \tau_i P(t, T_{i+1}).$$

Later we will be interested in swaps and their value functions. For now, suppose that we have the following equation:

$$cA_{k,m}(t) + P(t, T_{k+m}) = P(t, T_k).$$

If we solve this for c, we get the so-called swap or par rate. This is the only value of a constant fixed coupon at which the swap is worth exactly 0 at time t. The fixed coupon of the swap is usually set to the swap rate corresponding to the start date and tenor of the swap to ensure fair pricing.

Definition. A *swap rate* or *par rate* starting at T_k with tenor m is the rate $S_{k,m}$ whose value at time t is

$$S_{k,m}(t) = \frac{P(t, T_k) - P(t, T_{k+m})}{A_{k,m}(t)}.$$

Chapter 3

Interest Rate Derivatives

Interest rates is one of the most diverse asset classes. Since interest rates themselves are not tradable, it contains mainly interest rate derivatives of various kinds. The most simple and perhaps important one is the zero coupon bond. Almost all other derivatives incorporate it, as we will see, so its price dynamics is crucial to price other products. Now let us look at other quite popular interest rate derivatives: swaps and swaptions.

3.1 Swaps

Suppose we want to pay a fixed rate on our loan, but the terms are more advantageous if we take a floating rate. It is fair to assume that there exists someone in the exact opposite position. To everyone's benefit, we can write a contract and swap the loan payments so that we all end up paying according to our preferences. The product that allows this is called a swap.

Definition. A *swap* is a contract in which the two parties agree to swap two cash flows over a predefined time period with predefined terms. One of the parties pays a cash flow depending on a set of fixed rates, called the *fixed leg*, while the other pays a cash flow depending on a set of floating rates, called the *floating leg*.

The swap is called *payer swap* or *receiver swap*, depending on whether we – the first party – pay or receive the fixed leg, respectively. The length of the payment period is called the *tenor* of the swap. The payments are called *coupons*, and do not have to be constant throughout the period. The fixed and floating payment schedules do not have to be exactly aligned (but to every fixed payment there must correspond a floating one) and their frequencies usually do not match because of currency conventions. The floating

leg is usually tied to a reference rate with a spread correction. The two legs of the swap can be denominated in different currencies, which is called a cross-currency swap. Here, we cover only same-currency swaps. The notionals can also vary over time, and these constructions are very popular among those who want to hedge mortgage repayments, for example. A swap whose notionals increase over time is called an *accreter*, while decreasing notionals correspond to an *amortiser*.

It is clear that swaps are very diverse and we shall narrow our focus to show some important properties of them. For simplicity, we assume swaps with no spread, only one currency, and the same payment schedule for both legs. It is very simple to adjust calculations with the relaxation of the last assumption, however it would be confusing to introduce more notation and lengthy formulae and would hinder the understanding of the core concepts, which are not violated by this assumption anyway. Now we present the value of a swap:

Notation. The time $t \leq T_0$ value of a payer swap starting at T_0 with payment dates $T_i, \quad i = 1, \dots n$, fixed coupon schedule c_i, and notionals N_i is

$$V_{swap}(t) = \sum_{i=0}^{n-1} N_i \tau_i P(t, T_{i+1}) (L_i(t) - c_i)$$

where $\tau_i = T_{i+1} - T_i$ and $L_i(t) = L(t, T_i, T_{i+1})$ is the LIBOR rate.

Remark. Since the LIBOR rate can be rewritten as a function of the zero coupon bond, we obtain

$$\begin{aligned} V_{swap}(t) &= \sum_{i=0}^{n-1} N_i \tau_i P(t, T_{i+1}) (L_i(t) - c_i) = \\ &= \sum_{i=0}^{n-1} N_i P(t, T_{i+1}) \left(\frac{1}{P(t, T_i, T_{i+1})} - 1 - \tau_i c_i \right) = \\ &= \sum_{i=0}^{n-1} N_i P(t, T_{i+1}) \left(\frac{P(t, T_i)}{P(t, T_{i+1})} - 1 - \tau_i c_i \right) = \\ &= \sum_{i=0}^{n-1} N_i \left(P(t, T_i) - P(t, T_{i+1}) - \tau_i c_i P(t, T_{i+1}) \right). \end{aligned}$$

So-called vanilla swaps are very simple with unit notionals (or flat notional term structure which is essentially the same) and a flat fixed coupon term structure.

Remark. The value of a vanilla payer swap at time $t \leq T_0$ is

$$\begin{aligned} V_{vanilla}(t) &= \sum_{i=0}^{n-1} \left(P(t,T_i) - P(t,T_{i+1}) - c\tau_i P(t,T_{i+1})\right) = \\ &= \sum_{i=0}^{n-1} \left(P(t,T_i) - P(t,T_{i+1}) - c\tau_i P(t,T_{i+1})\right) = \\ &= P(t,T_0) - P(t,T_n) - c\sum_{i=0}^{n-1} \tau_i P(t,T_{i+1}). \end{aligned}$$

In the last step we used that the first part of the sum is telescopic. We recall the definition of the annuity and the swap rate and rearrange the equation.

$$\begin{aligned} V_{vanilla}(t) &= P(t,T_0) - P(t,T_n) - c\sum_{i=0}^{n-1} \tau_i P(t,T_{i+1}) = \\ &= \sum_{i=0}^{n-1} \tau_i P(t,T_{i+1}) \left(\frac{P(t,T_0) - P(t,T_n)}{\sum_{i=0}^{n-1} \tau_i P(t,T_{i+1})} - c\frac{\sum_{i=0}^{n-1} \tau_i P(t,T_{i+1})}{\sum_{i=0}^{n-1} \tau_i P(t,T_{i+1})} \right) = \\ &= A_{0,n}(t)\left(S_{0,n}(t) - c\right) = A_n(t)\left(S_n(t) - c\right). \end{aligned}$$

Now we see why the swap rate is also called the par rate: it is the rate of the fixed coupon at which the swap is fairly priced, in other words, its value is 0.

Notation. We can refine this value of vanilla swaps for more general swaps with notionals, but first, let $\delta N_i = N_i - N_{i+1}$ for $i = 0, \ldots, n-1$ and let $N_n \coloneqq 0$. This quantity is the difference in consecutive notionals. For accreters, it is negative; for amortisers, it is positive. For the vanilla swap, $\delta N_i = 0$ for $i = 0, \ldots, n-2$, and $\delta N_{n-1} = N_{n-1}$. Note that in every case $\sum_{i=0}^{n-1} \delta N_i = N_0 - N_n = N_0$.

Remark. The value of a swap with a general term structure of notionals, but a flat fixed coupon term structure, is

$$\begin{aligned} V_{swap}(t) &= \sum_{i=0}^{n-1} N_i \left(P(t,T_i) - P(t,T_{i+1}) - \tau_i c P(t,T_{i+1})\right) = \\ &= N_0 P(t,T_0) - \sum_{i=0}^{n-1} \delta N_i P(t,T_{i+1}) - c\sum_{i=0}^{n-1} N_i \tau_i P(t,T_{i+1}) = \end{aligned}$$

$$= N_0 P(t, T_0) - \sum_{i=0}^{n-1} (\delta N_i + c N_i \tau_i) P(t, T_{i+1}). \tag{3.1}$$

For vanilla swaps, it equals the previous value.

To make this formula more similar to that of vanilla swaps, we introduce a modified annuity factor and swap rate for general swaps. For the rest of the work, we will only cover swaps with a flat fixed coupon term structure, so henceforth we assume that $c_i = c$ for $i = 0, \ldots, n-1$.

Definition. The time t value of the $\tilde{A}_{k,m}$ *annuity* of a non-vanilla swap with start date T_k and tenor m is

$$\tilde{A}_{k,m}(t) = \sum_{i=k}^{k+m-1} N_i \tau_i P(t, T_{i+1}).$$

Definition. The time t value of the $\tilde{S}_{k,m}$ *swap rate* of a non-vanilla swap with start date T_k and tenor m is

$$\tilde{S}_{k,m}(t) = \frac{\sum_{i=k}^{k+m-1} N_i \left(P(t, T_i) - P(t, T_{i+1})\right)}{\tilde{A}_{k,m}(t)}.$$

Remark. Using the telescopic sum and the fact that $N_i = 1, \quad \forall i = 0, \ldots, n-1$, it is clear that these new definitions return the values previously defined for vanilla swaps.

Remark. We can rewrite the value of the swap as follows:

$$\begin{aligned}
V_{swap}(t) &= \sum_{i=0}^{n-1} N_i \left(P(t, T_i) - P(t, T_{i+1}) - \tau_i c P(t, T_{i+1})\right) = \\
&= \sum_{i=0}^{n-1} N_i \left(P(t, T_i) - P(t, T_{i+1})\right) - c \sum_{i=0}^{n-1} N_i \tau_i P(t, T_{i+1}) = \\
&= \sum_{i=0}^{n-1} N_i \tau_i P(t, T_{i+1}) \left(\frac{\sum_{i=0}^{n-1} N_i \left(P(t, T_i) - P(t, T_{i+1})\right)}{\sum_{i=0}^{n-1} N_i \tau_i P(t, T_{i+1})} - c \frac{\sum_{i=0}^{n-1} N_i \tau_i P(t, T_{i+1})}{\sum_{i=0}^{n-1} N_i \tau_i P(t, T_{i+1})} \right) = \\
&= \tilde{A}_{0,n}(t) \left(\tilde{S}_{0,n}(t) - c \right) = \tilde{A}_n(t) \left(\tilde{S}_n(t) - c \right).
\end{aligned}$$

It is a true extension of the formula for vanillas.

3.2 Swaptions

It is clear from the previous section that a swap can take both positive and negative values. Suppose that for some reason, for example extreme risk–aversion, we do not want to enter the swap if it is negative. Obviously, this only makes sense if the start date of the swap is in the future, hence the value is not yet known. The product which facilitates this – i.e. gives the *right but not the obligation* to enter into a contract – is called, under more general settings, an *option*, and the object of the purchase or sell is the *underlying*. Specifically, if the underlying is a swap, we talk about *swaptions*.

Remark. Generally, the option pay-off functions can be written in the form

$$V_{option}(T) = (h(U(T) - K))^{+} \tag{3.2}$$

where T is the *exercise date* in the future, when the holder can decide whether to enter into the contract, $U(T)$ denotes the underlying's price at T, K is the so-called *strike* or *exercise price*, agreed upon writing the contract, that is strictly before T, and h specifies the type of the contract: $h = 1$ for *call options*, where the holder has the right to buy at K and $h = -1$ for *put options*, where the right is to sell.

Of course, as usual in asset pricing, we get the time $t \leq T$ value of the discounted option – i.e. a martingale – by taking expectation under the risk-neutral measure:

$$V_{option}(t) = \beta(t)\,\mathbb{E}_t\left[\beta(T)^{-1} V_{option}(T)\right] = \beta(t)\,\mathbb{E}_t\left[\beta(T)^{-1}(h(U(T) - K))^{+}\right]$$

where β denotes the numéraire.

Remark. It follows from the reasoning above that the fair price of the swaption must be the expected discounted value of the positive part of the swap. It remains to be seen that it is truly an option, in other words, it can be written in a form similar to (3.2).

Recall that a payer swap's value at the start of the swap is given by

$$V_{swap}(T_0) = \tilde{A}_n(T_0)\left(\tilde{S}_n(T_0) - c\right).$$

It seems natural to choose c, that is, the fixed coupon as the strike price, the (generalised) swap rate as the underlying, and the (generalised) annuity as the numéraire. While the above formula remains unchanged for different dates t, we have to establish that t, that is, the exercise date of the swaption, must not be greater than the start date of the swap.

While it is possible to use a date *before* the start date, it complicates things, as it requires not only an initial yield curve known at the time of purchase, but assumptions about future yield curves, specifically at time t. Models which accommodate to these kinds of swaptions are called *mid-curve* models, and are wide-spread in industry, nonetheless, we will restrain our focus on swaptions whose exercise date coincides with the start date of the underlying. To conclude, we write the swaption's value in the following form:

Definition. The time $t \leq T_0$ value of a payer swaption with exercise date T_0 is

$$V_{swap}(t) = \tilde{A}_n(t)\,\mathbb{E}_t^{\tilde{A}}\left[\tilde{A}_n(T_0)^{-1}\,\tilde{A}_n(T_0)\left(\tilde{S}_n(T_0) - c\right)^+\right] = \tilde{A}_n(t)\,\mathbb{E}_t^{\tilde{A}}\left[\left(\tilde{S}_n(T_0) - c\right)^+\right]$$

where $\mathbb{E}_t^{\tilde{A}}$ denotes the expectation taken under the risk-neutral annuity measure, i.e. the measure under which the annuity is a martingale.

Remark. It is evident that the payer swaption is thus a call option of the swap rate, with the fixed coupon as strike. Similarly, the receiver swaption, i.e. the swaption of a receiver swap is a put option.

Remark. The previous examples were all European swaptions, with one predefined exercise date. On the market, however, path-dependent swaptions with early-exercise options are also very common and they require slightly different pricing techniques. The most important example is Bermudan swaptions with a predefined set of possible exercise dates. Bermudans are more complex to price and there is fast-expanding literature on how to do this as precisely as possible[1].

[1] See e.g. Gatarek & Jabłecki (2021) or Hagan (2000).

Chapter 4

Swaption Pricing Techniques

4.1 The Jamshidian Decomposition

Recall that the swaption's value at the exercise date T_0 is given by

$$V_{swaption}\left(T_0\right) = \left(N_0 P\left(T_0, T_0\right) - \sum_{i=0}^{n-1} \left(\delta N_i + c N_i \tau_i\right) P\left(T_0, T_{i+1}\right) \right)^+, \tag{4.1}$$

which means that it is essentially an option – the positive part – of a sum, independently of model choice. A typical method to compute swaption prices – which will be discussed in details later on in Section 4.3 – is by Monte Carlo: simulating multiple paths for the interest rates and calculating the value of the swaption for each of them, then averaging across the scenarios. Based on the model framework we operate in, this can be cumbersome and, for a couple of popular model choices, not the most efficient. In several cases, an analytical solution exists for the zero coupon bond price, and here we present a trick which makes use of that result.

Suppose we have an interest rate model in which there exists a stochastic process $x\left(t\right)$ satisfying the following conditions:

1. x is a deterministic function of an interest rate variable whose stochastic differential equation is known to us, in other words, we transfer all randomness into x and switch to its differential equation as the driving dynamics of the system.

2. There exists a bond reconstitution formula, i.e. a formula for the zero coupon bond $P\left(t, T\right)$ for any $t \leq T$ in the present or future, which is a deterministic and **monotone** function of $x\left(t\right)$.

Now we look at the swaption price again at T_0 and to accentuate the dependence of the zero coupon bond price on $x(T_0)$, write $P(T_0, T_i, x(T_0)) \coloneqq P(T_0, T_i)$ for $i = 1, \ldots, n$:

$$V_{swaption}(T_0) = \left(N_0 - \sum_{i=0}^{n-1} (\delta N_i + cN_i\tau_i) P(T_0, T_{i+1}, x(T_0)) \right)^+ .$$

We use Jamshidian (1989)'s trick and find a value x^* for which the swap is worth exactly zero at the exercise date:

$$N_0 = \sum_{i=0}^{n-1} (\delta N_i + cN_i\tau_i) P(T_0, T_{i+1}, x^*) . \tag{4.2}$$

If we have a well-behaving model and thus the zero coupon bond price can range from about 0 to 1 (or, assuming negative short rates, even greater), then such a value exists and, by monotonicity of $P(t, T, x(t))$ in $x(t)$, is unique. This gives us a well-defined solution.

We define the so-called *strikes*[1] as follows:

$$K_i \coloneqq P(T_0, T_i, x^*) \tag{4.3}$$

and thus, by the definition of x^*, we have:

$$N_0 = \sum_{i=0}^{n-1} (\delta N_i + cN_i\tau_i) K_{i+1}. \tag{4.4}$$

Without loss of generality, suppose that the price of the bond is monotonically decreasing, as is the case with most short rate models, in x. Now this means that the swaption is worth a positive value if and only if $x(T_0) > x^*$. We insert (4.4) into the swaption price equation to derive a simpler formula:

$$\begin{aligned} V_{swaption}(T_0) &= \left(N_0 - \sum_{i=0}^{n-1} (\delta N_i + cN_i\tau_i) P(T_0, T_{i+1}, x(T_0)) \right)^+ = \\ &= \left(\sum_{i=0}^{n-1} (\delta N_i + cN_i\tau_i) K_{i+1} - \sum_{i=0}^{n-1} (\delta N_i + cN_i\tau_i) P(T_0, T_{i+1}, x(T_0)) \right)^+ = \end{aligned}$$

[1] As we will shortly see, they are named this for a reason.

$$= \left(\sum_{i=0}^{n-1} \left(\delta N_i + c N_i \tau_i \right) \left(K_{i+1} - P\left(T_0, T_{i+1}, x\left(T_0 \right) \right) \right) \right) \chi_{x(T_0) > x^*} =$$
$$= \sum_{i=0}^{n-1} \left(\delta N_i + c N_i \tau_i \right) \left(K_{i+1} - P\left(T_0, T_{i+1}, x\left(T_0 \right) \right) \right)^+ . \tag{4.5}$$

We used that x^* is the only value for which $K_i = P\left(T_0, T_i, x^*\right)$, and that these bond prices are monotone in $x\left(T_0\right)$.

With this little trick, we decomposed the swaption – an option on a sum – into a sum of put options on zero coupon bonds with strikes K_i, which are in most cases easier to price, for instance, with the Black formula or simpler Monte Carlo. Notice that the same reasoning works for receiver swaptions, and the result will be a basket of call options.

Note that the decomposition is model-independent and only requires two rather easy-to-check conditions for the zero coupon bond price. As a consequence, the Jamshidian decomposition could be widely used for pricing swaptions, and seems commendable since it gives an analytical result. In practice, however, it is rarely put to use. Although it is analytical, it requires a root search, which alone can result in numerical errors, and pricing of a basket of put options. Hence while it is a perfectly fine method for some models, e.g. the Gaussian short rate model, it is completely forgotten in other settings.

We will explore this method in three models, where there exists an explicit formula for bond option prices as well: the Vasicek (1977), the Hull & White (1990) and the Cox et al. (1985).

4.2 Tree Methods

One problem with an analytical approach is that in several cases it simply does not exist. If there is no analytical zero coupon bond price as in, for instance, the Black & Karasinski (1991) model, or the model is not Markovian, e.g. in a non-separable Heath–Jarrow–Morton (Heath et al. (1992)) setting, there is no hope to price a swaption analytically. Even if there exists an analytical formula for the bond price, but does not exist for the bond *option* price, we are not much further with the Jamshidian trick, since we still have to price a basket of options to get the swaption price right. In these cases it is inevitable to use some sort of numerical approximation.

In this dissertation, we will cover two of the most well-spread techniques: tree methods and Monte Carlo simulation. While both of them make assumptions, and naturally, as

all numerical methods, use a discretisation scheme, the two differ in several aspects. Tree methods, much like finite difference methods[2], work on a finite – tree-like – grid, the horizontal axis representing discretised time, and the vertical one possible states at each of these time points. The tree method makes assumptions – derived from the SDE – about how the state variable possibly evolves over time on this grid, and working backwards from the end, where the price of the instrument is known, takes expectation, discounts the value in each time step, and finally prices the instrument at the origin.

It is clear from this setting that trees are able to price so-called path-dependent instruments, e.g. American, Bermudan, Asian, look-back, knockout, or digital options, as opposed to classical Monte Carlo, which can only handle the simple European case. Monte Carlo methods generate multiple trajectories of the state variable, price the instrument in each of these, take the mean and discount to get the price at the origin. While constructing and calibrating a tree can be difficult, the work put into setting up the engine pays-off runtime-wise. The tree method does *not* include any randomness, it is only necessary to run it once, while generating paths in the simulation is often time- and memory-consuming, especially if we want precise results and convergence is slow. The tree method, however, is very limited in what scenarios can play out. It essentially wipes out the possibility of tail events happening early on and concentrates on the middle of the distribution. There are several examples where it is worth switching to Monte Carlo, but usually, when the model allows, industry employs both methods, even in cases where an analytical solution could be found. We will do the same and first concentrate on tree methods.

The wide-spread use of trees in financial applications was probably triggered by the seminal work of Cox et al. (1979), mere years after the Black–Scholes–Merton (BSM) model (Black & Scholes (1973) and Merton (1973)) model was published. The so-called Cox–Ross–Rubinstein model or, in short, the CRR model develops a binomial tree to price options. It works under very similar assumptions as the BSM model, and it was shown that, by decreasing the time step, the price returned by the CRR tree converges to the BSM price[3].

Binomial trees, in general, use a finite time grid with usually but not necessarily equidis-

[2]We will not cover these in this dissertation, but the idea is to solve the SDE or PDE on a finite grid, working usually backwards from boundary conditions and using discretisation. It is a highly efficient method in most cases and is widely used in industry and research. The most famous technique is the Crank & Nicolson (1947) scheme, but for a comprehensive summary, refer to, e.g., Andersen & Piterbarg (2010), pp. 43-93.

[3]See e.g. Pascucci (2011) Section 2.3.5.

tant points. The instrument starts from a deterministic value (in other words, adapted to the filtration at time 0) at the origin and in each time step it moves up or down by a predefined amount or rate. Since this evolution can take only two values, the tree is called binomial.

Without making any further assumptions, a tree of n steps can have 2^n nodes in the nth step. Since this exponentiality can hinder any run time advantage coming from precluding randomness, one usually constructs the tree so that it grows linearly: one additional node a step, instead of doubling. This can be achieved by a so-called *recombining tree*, where the values are defined so that an up-down and a down-up scenario in two steps starting from the same node ends at the same node as well. In this way, in the nth step, we have only $n + 1$ nodes, making the calculations much faster. In addition to the up and down values, the other defining characteristic of a binomial tree is the *probability* of moving up from each node. Note that this must not necessarily be the same for every node.

Specific models differ in how they define these values and probabilities. However, though there are a multitude of ways to construct the tree, it cannot price efficiently if the parameters are not calibrated so that movements on the tree mimic the dynamics of the instrument. In other words, one must ensure, by adjusting parameters, that one captures the characteristics of the distribution at each time step. Since we usually work in continuous-time models and trees are inherently discrete, there is no chance we can replicate *all* attributes of a distribution. We must make sacrifices, and most of the time tree developers are already satisfied if some basic metrics of the distribution are matched. The best-known technique is *moment matching*[4], where the parameters are calibrated so that some of the first moments – or central moments – of the distribution are fit perfectly. This usually comprises of the mean and the variance, and in more elaborate schemes, perhaps the skew or kurtosis.

In the binomial model, since we have three parameters – up value, down value, up probability – and one constraint is imposed by the recombining characteristic of the tree, we can only hope to match the mean and variance. The CRR mentioned above is an equity option model, where the underlying is assumed to follow a log-normal distribution, so the up and down parameters are coefficients of the previous price and are reciprocals of each other, namely $e^{\sigma\sqrt{\Delta t}}$ and $e^{-\sigma\sqrt{\Delta t}}$. The probability is computed so that the mean matches the risk-free rate of return, in accordance with the Black–Scholes and general no-arbitrage theory world. In other settings, the probability is assumed to be $\frac{1}{2}$ and the up value is

[4]The method was originally introduced by Chebyshev.

adjusted to match the mean.

When pricing swaptions, our state variable is usually an interest rate, most often the short rate. Although there exist some models, e.g., the Black et al. (1990) and the Black & Karasinski (1991) where the short rate is log-normal, most often it is not, and thus an additive tree is a more appropriate choice than a multiplicative one like the CRR. In particular, all three models covered in this dissertation are better suited for the additive version, since the Vasicek and Hull–White models are normal, while the CIR is non-central chi-squared. However, it is important to note in advance that the Hull–White model is different in the sense that it is a no-arbitrage model, while the others are equilibrium models, meaning that it can fit the whole yield curve. Naturally, these kind of models are more likely employed in industry due to their ability to reproduce market data. Other no-arbitrage models include the Ho & Lee (1986), the Black & Karasinski (1991) and the Black–Derman–Toy (Black et al. (1990)).

They attract more attention and research, and the existence of a time-dependent mean requires that the mean itself be represented in the tree. Since in binomial models, in every second time step, the mean is always between an up and down value, the development of trinomial – recombining – trees followed. Hull and White themselves were the first to propose a trinomial tree for their extended Vasicek model (Hull & White (1993)), which they later refined in Hull (1996) and extended to other no-arbitrage models like the Black & Karasinski (1991) in Hull & White (1994*b*) and multi-factor models in Hull & White (1994*a*). While for no-arbitrage models it is desirable to keep the mean in the tree in every time step, in equilibrium models, which are not very often used for actual pricing purposes, only theoretical research, and maybe long-term prediction, the binomial model is perfectly adequate. Therefore, we will introduce a binomial lattice tree model using Nelson & Ramaswamy (1990) for the Vasicek and CIR models in Sections 5.3.3 and 5.6.4, and present the Hull & White (1993) trinomial tree in Section 5.5.4.

4.3 Monte Carlo Simulation

The most wide-spread method to price instruments numerically is known as Monte Carlo simulation and is often used in industry and academic settings. Unlike tree methods, it centres around the idea of randomness, so it can give a better approximation of the distribution, but its inability to price path-dependent instruments shadows its merits.[5]

[5]Though generally this is true, there have been several attempts to develop e.g. American Monte Carlo-s by introducing alterations to make them suitable for path-dependencies. For general theories on

Nonetheless, it is a generally accepted pricing technique, and frequently used as a benchmarking tool to test the efficiency of other newly-developed algorithms.

The key idea behind Monte Carlo is to generate a family of possible paths – trajectories – of the data-generating process, calculate the quantity in question in each scenario, then take expectation and calculate the present value of the average. It is based on the very well known "expected present value" principle of asset pricing, in other words, the theorem that the discounted process is a martingale under the risk-neutral measure.

While it seems very straightforward, each of the above steps poses several problems. Firstly, though generating paths appears to be an easy task – simply follow the recipe, that is, the SDE in question – it is not at all obvious in continuous time. Taking expectation is also problematic, partly because the variance can be high enough to make such calculations quite volatile and imprecise. Finally, whilst discounting seems to be the easiest part, it always includes the choice of the numéraire, and by consequence, the risk neutral measure, which affects both previous steps as well. For the purpose of pricing swaptions, most often we operate with the $P(0,t)$ zero coupon bond price or the annuity as numéraire. Below, based on Kloeden & Platen (1992) and Andersen & Piterbarg (2010)'s works, we discuss several ways to generate trajectories.[6]

4.3.1 (Explicit) Euler–Maruyama Scheme

Suppose we have a stochastic differential equation (SDE) for a variable x of the form:

$$dx(t) = \mu(t,x)\,dt + \sigma(t,x)\,dW(t) \tag{4.6}$$

where $W(t)$ is a Wiener process in its natural filtration.

We want to generate paths of x starting from a known initial value $x_0 \coloneqq x(0)$. If time is discrete, we can simply use each point on the time grid to evaluate x. However, if time is continuous, we have to solve a discretisation problem. We choose a suitable set of points $\{t_i\}_{i=0}^{n}$ and treat the SDE as if it were a difference equation on the aforementioned time grid. If it is fine enough – and μ and σ satisfy some conditions[7] – then the discretised simulation can be a good approximation of the continuous trajectory. There exist a variety of methods to make use of the t_i-s in the discretised SDE, amongst which

this topic, refer to Rogers (2002), for application to Bermudan swaptions, see e.g. Leclerc et al. (2009).

[6] For variance reduction techniques see e.g. Andersen & Piterbarg (2010) pp. 140-155. or Kloeden & Platen (1992) pp. 511-528.

[7] They must satisfy a uniform Lipschitz condition and a growth bound (Kloeden & Platen (1992)).

the most straightforward and best known is the Euler–Maruyama scheme.

It is a first-order approximation, and only uses past values of x or t in μ or σ to calculate the next x. Its general form is as follows:

$$x_{i+1} - x_i = \mu(t_i, x_i)\,\Delta_i + \sigma(t_i, x_i)\,\sqrt{\Delta_i} Z_i$$

where $\Delta_i \coloneqq t_{i+1} - t_i$ and Z_i is a standard Gaussian variable, $\{Z_i\}_{i=0}^{n-1}$ being independent. Here we used the fact that the increments of the Wiener process are independent and Gaussian with mean 0 and variance the length of the interval.

After a small rearrangement, we get

$$x_{i+1} = x_i + \mu(t_i, x_i)\,\Delta_i + \sigma(t_i, x_i)\,\sqrt{\Delta_i} Z_i. \tag{4.7}$$

In the very core of the Monte Carlo simulation is the Law of Large Numbers, which states that if we have identically distributed independent random variables $\{x_i\}_{i=1}^{n}$ with finite mean $\mathbb{E}[x]$ then

$$\frac{\sum_{i=1}^{n} x_i}{n} \underset{(n\to\infty)}{\longrightarrow} \mathbb{E}[x].$$

So it is obvious that we want our x-s to converge, otherwise the method is useless. If we inspect (4.7) – without further submerging into definitions and rates of weak or strong convergence[8] – we note that the speed of convergence in the drift and diffusion term is not the same. Indeed, the latter converges more slowly, caused by the square root. This is one major drawback of the Euler–Maruyama scheme and will be dealt with in higher-order schemes such as the Milstein scheme.

Another serious issue arises when μ takes certain forms. In particular, let $\mu(t, x) \coloneqq \alpha x(t)$, that is, linear in x, as is the case with all Ornstein–Uhlenbeck-type mean reverting processes. Now (4.7) looks like

$$x_{i+1} = x_i(1 + \alpha\Delta_i) + \sigma(t_i, x_i)\,\sqrt{\Delta_i} Z_i$$

Again, without going into details about the exact meaning of stability[9], we shall note that this system is stable only if $|1 + \alpha\Delta_i| < 1$, which practically means $\alpha\Delta_i > -2$. It

[8] Kloeden & Platen (1992) pp. 323-330.
[9] Kloeden & Platen (1992) pp. 331-338.

is obvious then that there shall be restrictions on the length of the time step to ensure stability. To avoid this kind of restriction, which is highly inconvenient when pricing long-dated products, for instance, we can revisit the – explicit – Euler–Maruyama scheme and introduce little changes to prevent instability and thus accommodate for a wider range of μ coefficients.

4.3.2 Linear Drift Euler Scheme

Suppose we have a mean reverting SDE of the form:

$$dX(t) = \varkappa(\vartheta(t) - X(t))\,dt + \sigma(t, X(t))\,dW(t).$$

Recall from the Euler–Maruyama scheme that this is a linear drift process, so unless we impose restrictions on the time step, we might end up with an unstable system. Fortunately, in this special case, this problem can be solved by change of variables (Andersen & Piterbarg (2010)). Let

$$y(t) = e^{\varkappa t}X(t) - \varkappa\int_0^t e^{\varkappa u}\vartheta(u)\,du.$$

By Ito's lemma:

$$\begin{aligned}
dy(t) &= \varkappa e^{\varkappa t}X(t) - \varkappa e^{\varkappa t}\vartheta(t)\,dt + e^{\varkappa t}\,dX(t) + \frac{1}{2}0 = \\
&= \varkappa e^{\varkappa t}X(t) - \varkappa e^{\varkappa t}\vartheta(t) - e^{\varkappa t}\varkappa(\vartheta(t) - X(t))\,dt + e^{\varkappa t}\sigma(t, X(t))\,dW(t) = \\
&= e^{\varkappa t}\sigma(t, X(t))\,dW(t) = \\
&= e^{\varkappa t}\sigma\left(t, e^{-\varkappa t}\left(y(t) + \varkappa\int_0^t e^{\varkappa u}\vartheta(u)\,du\right)\right)dW(t). \qquad (4.8)
\end{aligned}$$

It can be proved (Kloeden & Platen (1992)) that the Euler–Maruyama scheme for $y(t)$ is stable. We will use (4.8) for a variety of models including the Vasicek, the one-factor Hull–White and the Cox–Ingersoll–Ross.

4.3.3 Implicit Euler–Maruyama Scheme

Another – more general – trick to stabilise the (explicit) Euler–Maruyama discretisation is to substitute into the drift term at the *end* of each interval, instead of the beginning. This

is called *implicit Euler–Maruyama scheme*, as opposed to its original form, the explicit one. Using the notation established at the beginning of the section, it translates into[10]

$$x_{i+1} = x_i + \mu\left(t_{i+1}, x_{i+1}\right)\Delta_i + \sigma\left(t_i, x_i\right)\sqrt{\Delta_i}Z_i. \tag{4.9}$$

Let us revisit the simplest linear case:

$$dx\left(t\right) = \alpha x\left(t\right)dt + \sigma\left(t, x\left(t\right)\right)dW\left(t\right).$$

By the implicit Euler scheme we get:

$$\begin{aligned} x_{i+1} &= x_i + \alpha x_{i+1}\Delta_i + \sigma\left(t_i, x_i\right)\sqrt{\Delta_i}Z_i \\ &= \frac{1}{1-\alpha\Delta_i}\left(x_i + \sigma\left(t_i, x_i\right)\sqrt{\Delta_i}Z_i\right). \end{aligned} \tag{4.10}$$

It is known from ODE-s[11] that this system is stable if and only if $|\frac{1}{1-\alpha\Delta_i}| < 1$ and $\alpha\left(\in\mathbb{R}\right) < 0$, which is easily satisfied.

Finally, in the important case of the mean reverting process, the implicit scheme is as follows:

$$x_{i+1} = \frac{1}{1+\varkappa\left(t_{i+1}\right)\Delta_i}\left(x_i + \varkappa\left(t_{i+1}\right)\vartheta\left(t_{i+1}\right)\Delta_i + \sigma\left(t_i, x_i\right)\sqrt{\Delta_i}Z_i\right). \tag{4.11}$$

Though it is an important result, we do not consider implicit schemes in this dissertation.

4.3.4 Milstein Scheme

Recall that one problem with the explicit Euler–Maruyama scheme was that the drift and diffusion terms converged at different speeds, namely the former linearly while the latter at $\mathcal{O}\left(\sqrt{\Delta}\right)$. Our aim is to "accelerate" the diffusion term. To achieve this, we follow Milshtein (1979) and use an Ito–Taylor expansion.

Apply Ito's lemma to $\mu\left(u, x\left(u\right)\right)$:

$$\mu\left(u, x\left(u\right)\right) = \mu\left(t_i, x\left(t_i\right)\right) + \int_{t_i}^{u}\frac{\partial}{\partial s}\mu\left(s, x\left(s\right)\right)ds + \int_{t_i}^{u}\frac{\partial}{\partial x}\mu\left(s, x\left(s\right)\right)dx\left(s\right) +$$

[10] Notice that only the drift term is evaluated at the end of the interval. The same cannot be done for the volatility term; for exact reasons, see Kloeden & Platen (1992).

[11] Kloeden & Platen (1992) pp. 292-298.

$$
\begin{aligned}
&+\frac{1}{2}\int_{t_i}^{u}\frac{\partial^2}{\partial x^2}\mu\left(s,x\left(s\right)\right)d\left[x\right]=\\
&=\mu\left(t_i,x\left(t_i\right)\right)+\int_{t_i}^{u}\frac{\partial}{\partial s}\mu\left(s,x\left(s\right)\right)ds+\int_{t_i}^{u}\mu\left(s,x\left(s\right)\right)\frac{\partial}{\partial x}\mu\left(s,x\left(s\right)\right)ds+\\
&+\int_{t_i}^{u}\sigma\left(s,x\left(s\right)\right)\frac{\partial}{\partial x}\mu\left(s,x\left(s\right)\right)dW\left(s\right)+\frac{1}{2}\int_{t_i}^{u}\sigma\left(s,x\left(s\right)\right)^2\frac{\partial^2}{\partial x^2}\mu\left(s,x\left(s\right)\right)ds=\\
&=\mu\left(t_i,x\left(t_i\right)\right)+\int_{t_i}^{u}\left(\frac{\partial}{\partial s}+\mu\left(s,x\left(s\right)\right)\frac{\partial}{\partial x}+\frac{1}{2}\sigma\left(s,x\left(s\right)\right)^2\frac{\partial^2}{\partial x^2}\right)\mu\left(s,x\left(s\right)\right)ds+\\
&+\int_{t_i}^{u}\sigma\left(s,x\left(s\right)\right)\frac{\partial}{\partial x}\mu\left(s,x\left(s\right)\right)dW\left(s\right)=\\
&=:\mu\left(t_i,x\left(t_i\right)\right)+\int_{t_i}^{u}\mathcal{L}_0\mu\left(s,x\left(s\right)\right)ds+\int_{t_i}^{u}\mathcal{L}_1\mu\left(s,x\left(s\right)\right)dW\left(s\right)
\end{aligned}
\tag{4.12}
$$

where we introduce the operators

$$\mathcal{L}_0 \coloneqq \frac{\partial}{\partial s}+\mu\left(s,x\left(s\right)\right)\frac{\partial}{\partial x}+\frac{1}{2}\sigma\left(s,x\left(s\right)\right)^2\frac{\partial^2}{\partial x^2}$$

and

$$\mathcal{L}_1 \coloneqq \sigma\left(s,x\left(s\right)\right)\frac{\partial}{\partial x}.$$

Similarly, for $\sigma\left(u,x\left(u\right)\right)$ we get:

$$\sigma\left(u,x\left(u\right)\right)=\sigma\left(t_i,x\left(t_i\right)\right)+\int_{t_i}^{u}\mathcal{L}_0\sigma\left(s,x\left(s\right)\right)ds+\int_{t_i}^{u}\mathcal{L}_1\sigma\left(s,x\left(s\right)\right)dW\left(s\right). \tag{4.13}$$

By inserting these terms into the integral form of (4.6), we obtain:

$$
\begin{aligned}
x_{i+1}-x_i=&\int_{t_i}^{t_i+\Delta_i}\left(\mu\left(t_i,x\left(t_i\right)\right)+\int_{t_i}^{u}\mathcal{L}_0\mu\left(s,x\left(s\right)\right)ds+\int_{t_i}^{u}\mathcal{L}_1\mu\left(s,x\left(s\right)\right)dW\left(s\right)\right)du+\\
&+\int_{t_i}^{t_i+\Delta_i}\left(\sigma\left(t_i,x\left(t_i\right)\right)+\int_{t_i}^{u}\mathcal{L}_0\sigma\left(s,x\left(s\right)\right)ds+\int_{t_i}^{u}\mathcal{L}_1\sigma\left(s,x\left(s\right)\right)dW\left(s\right)\right)dW\left(u\right).
\end{aligned}
$$

Now apply the Euler–Maruyama scheme to μ and σ, that is, substitute into each partial derivative at the beginning of the interval, that is, at t_i:

$$x_{i+1} - x_i = \mu(t_i, x(t_i))\,\Delta_i + \mathcal{L}_0\mu(t_i, x(t_i)) \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} ds\,du + \mathcal{L}_1\mu(t_i, x(t_i)) \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} dW(s)\,du +$$

$$+ \sigma(t_i, x(t_i))\,(W(t_i+\Delta_i) - W(t_i)) + \mathcal{L}_0\sigma(t_i, x(t_i)) \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} ds\,dW(u) +$$

$$+ \mathcal{L}_1\sigma(t_i, x(t_i)) \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} dW(s)\,dW(u) =$$

$$=: \mu(t_i, x(t_i))\,\Delta_i + \sigma(t_i, x(t_i))\,(W(t_i+\Delta_i) - W(t_i)) +$$

$$+ \mathcal{L}_0\mu(t_i, x(t_i))\,I_{0,0} + \mathcal{L}_1\mu(t_i, x(t_i))\,I_{1,0} + \mathcal{L}_0\sigma(t_i, x(t_i))\,I_{0,1} + \mathcal{L}_1\sigma(t_i, x(t_i))\,I_{1,1}.$$

We have to compute the four double integrals. The first is a Riemann integral:

$$I_{0,0} \coloneqq \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} ds\,du = \int_{t_i}^{t_i+\Delta_i} u - t_i\,du = \left[\frac{(u-t_i)^2}{2}\right]_{t_i}^{t_i+\Delta_i} = \frac{\Delta_i^2}{2}.$$

The second:

$$I_{1,1} \coloneqq \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} dW(s)\,dW(u) = \int_{t_i}^{t_i+\Delta_i} W(u) - W(t_i)\,dW(u) =$$
$$= \frac{1}{2}W(t_i+\Delta_i)^2 - \frac{1}{2}W(t_i)^2 - \frac{1}{2}(t_i+\Delta_i) + \frac{1}{2}t_i - W(t_i)\,(W(t_i+\Delta_i) - W(t_i)) =$$
$$= \frac{1}{2}(W(t_i+\Delta_i) - W(t_i))^2 - \frac{1}{2}\Delta_i.$$

We used Ito's lemma to compute $\int W(u)\,dW(u) = \frac{1}{2}\left[W(u)^2 - \frac{1}{2}u\right]$. The third one:

$$I_{1,0} \coloneqq \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} dW(s)\,du = \int_{t_i}^{t_i+\Delta_i}\int_{s}^{t_i+\Delta_i} du\,dW(s) = \int_{t_i}^{t_i+\Delta_i} t_i + \Delta_i - s\,dW(s).$$

We used Fubini's theorem to change the order of integration[12]. This stochastic integral is known to be Gaussian[13] with mean 0. To compute the variance, we use the Ito-isometry[14]:

$$\mathbb{D}^2\left[I_{0,1}\right] = \int_{t_i}^{t_i+\Delta_i} \left(t_i+\Delta_i-s\right)^2 ds = \left[-\frac{\left(t_i+\Delta_i-s\right)^3}{3}\right]_{t_i}^{t_i+\Delta_i} = \frac{\Delta_i^3}{3}.$$

Finally, the last one:

$$\begin{aligned} I_{0,1} &:= \int_{t_i}^{t_i+\Delta_i}\int_{t_i}^{u} ds\, dW\left(u\right) = \int_{t_i}^{t_i+\Delta_i} u - t_i\, dW\left(u\right) = \int_{t_i}^{t_i+\Delta_i} u\, dW\left(u\right) - \int_{t_i}^{t_i+\Delta_i} t_i\, dW\left(u\right) \\ &= \int_{t_i}^{t_i+\Delta_i} u\, dW\left(u\right) - t_i\left(W\left(t_i+\Delta_i\right)\right). \end{aligned}$$

Integrating by parts, we get

$$\int_{t_i}^{t_i+\Delta_i} u\, dW\left(u\right) = \left[uW\left(u\right)\right]_{t_i}^{t_i+\Delta_i} - \int_{t_i}^{t_i+\Delta_i} W\left(u\right) du\,.$$

Using this result, we obtain:

$$\begin{aligned} I_{0,1} &= \left[uW\left(u\right)\right]_{t_i}^{t_i+\Delta_i} - \int_{t_i}^{t_i+\Delta_i} W\left(u\right) du - t_i\left(W\left(t_i+\Delta_i\right)\right) = \\ &= \Delta_i\left(W\left(t_i+\Delta_i\right) - W\left(t_i\right)\right) + \Delta_i W\left(t_i\right) - \int_{t_i}^{t_i+\Delta_i} W\left(u\right) du = \\ &= \Delta_i\left(W\left(t_i+\Delta_i\right) - W\left(t_i\right)\right) - \int_{t_i}^{t_i+\Delta_i} W\left(u\right) - W\left(t_i\right) du = \end{aligned}$$

[12]The integrand is bounded and predictable.

[13]It is, by definition of the Ito integral, and since the integrand is deterministic and adapted to the filtration, a linear combination of Gaussian variables with mean 0.

[14]Ito-isometry: if X is square-integrable and adapted to W's natural filtration, then

$$\mathbb{E}\left[\left(\int_0^T X\left(t\right) dW\left(t\right)\right)^2\right] = \mathbb{E}\left[\int_0^T X\left(t\right)^2 dt\right].$$

$$= \Delta_i \left(W\left(t_i + \Delta_i\right) - W\left(t_i\right) \right) - I_{1,0}.$$

We want to compute the covariance of $W\left(t_i + \Delta_i\right) - W\left(t_i\right)$ and $I_{1,0}$. By consequence of the Ito-isometry[15], and using the fact that $W\left(t_i + \Delta_i\right) - W\left(t_i\right) = \int_{t_i}^{t_i+\Delta_i} dW\left(u\right)$:

$$Cov\left[W\left(t_i + \Delta_i\right) - W\left(t_i\right), I_{1,0}\right] = \int_{t_i}^{t_i+\Delta_i} \left(t_i + \Delta_i - u\right) du = \left[-\frac{\left(t_i + \Delta_i - u\right)^2}{2} \right]_{t_i}^{t_i+\Delta_i} = \frac{\Delta_i^2}{2}.$$

So the variable $\begin{bmatrix} W\left(t_i + \Delta_i\right) - W\left(t_i\right) \\ I_{1,0} \end{bmatrix}$ is Gaussian with covariance matrix $\begin{bmatrix} \Delta_i & \frac{\Delta_i^2}{2} \\ \frac{\Delta_i^2}{2} & \frac{\Delta_i^3}{3} \end{bmatrix}$.

Thus the correlation is $\rho := \frac{Cov(\cdot)}{\sigma_1 \sigma_2} = \sqrt{\frac{3\Delta_i^4}{4\Delta_i \Delta_i^3}} = \sqrt{\frac{3}{4}}$.

Using the Cholesky decomposition[16] for the bi-variate case, we produce these two linearly dependent Gaussian variables as linear combinations of independent standard Gaussian variables $Y_{i,1}$ and $Y_{i,2}$:

$$\begin{bmatrix} W\left(t_i + \Delta_i\right) - W\left(t_i\right) \\ I_{1,0} \end{bmatrix} = \begin{bmatrix} \sqrt{\Delta_i} Y_{i,1} \\ \sqrt{\frac{\Delta_i^3}{3}} \sqrt{\frac{3}{4}} Y_{i,1} + \sqrt{\frac{\Delta_i^3}{3}} \sqrt{\frac{1}{4}} Y_{i,2}. \end{bmatrix}$$

Now all we have to do is insert everything into the discretised SDE to obtain the Milstein scheme:

$$\begin{aligned} x_{i+1} = x_i &+ \mu\left(t_i, x\left(t_i\right)\right) \Delta_i + \sigma\left(t_i, x\left(t_i\right)\right) \sqrt{\Delta_i} Y_{i,1} + \\ &+ \mathcal{L}_0 \mu\left(t_i, x\left(t_i\right)\right) \frac{\Delta_i^2}{2} + \end{aligned}$$

[15] If $I_1(t) = \int_0^t \alpha(s)\, dW(s)$ and $I_2(t) = \int_0^t \beta(s)\, dW(s)$, then (see e.g. Baldi (2017) p. 189.)

$$Cov\left[I_1(s), I_2(t)\right] = \int_0^{s \wedge t} \alpha(u)\, \beta(u)\, du\,.$$

[16] If a matrix $\mathbf{Q}$ is positive definite, there exists a unique lower-triangle matrix $\mathbf{S}$ so that $\mathbf{Q} = \mathbf{S}^\top \mathbf{S}$. In the specific case of a bi-variate correlation matrix with correlation ρ, $\mathbf{S} = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}$.

$$+ \mathcal{L}_1\mu\left(t_i, x\left(t_i\right)\right)\left(\sqrt{\frac{\Delta_i^3}{3}}\sqrt{\frac{3}{4}}Y_{i,1} + \sqrt{\frac{\Delta_i^3}{3}}\sqrt{\frac{1}{4}}Y_{i,2}\right) +$$
$$+ \mathcal{L}_0\sigma\left(t_i, x\left(t_i\right)\right)\left(\Delta_i Y_{i,1}\sqrt{\Delta_i} - \left(\sqrt{\frac{\Delta_i^3}{3}}\sqrt{\frac{3}{4}}Y_{i,1} + \sqrt{\frac{\Delta_i^3}{3}}\sqrt{\frac{1}{4}}Y_{i,2}\right)\right) +$$
$$+ \mathcal{L}_1\sigma\left(t_i, x\left(t_i\right)\right)\left(\frac{1}{2}Y_{i,1}^2\Delta_i - \frac{1}{2}\Delta_i\right). \tag{4.14}$$

Though it is useful to see the equation as a function of only deterministic terms and independent standard Gaussian variables, in reality, the generation of random variables precedes the generation of trajectories in the algorithm; therefore, it makes sense to rewrite (4.14) in a simpler form, with correlating standard Gaussian variables $Z_{i,1}$ and $Z_{i,2}$ where the correlation is $\sqrt{\frac{3}{4}}$.[17]

Our final formula for the Milstein scheme is thus:

$$x_{i+1} = x_i + \mu\left(t_i, x\left(t_i\right)\right)\Delta_i + \sigma\left(t_i, x\left(t_i\right)\right)\sqrt{\Delta_i}Z_{i,1} + \mathcal{L}_0\mu\left(t_i, x\left(t_i\right)\right)\frac{\Delta_i^2}{2} +$$
$$+ \mathcal{L}_1\mu\left(t_i, x\left(t_i\right)\right)\sqrt{\frac{\Delta_i^3}{3}}Z_{i,2} +$$
$$+ \mathcal{L}_0\sigma\left(t_i, x\left(t_i\right)\right)\left(\Delta_i Z_{i,1}\sqrt{\Delta_i} - \sqrt{\frac{\Delta_i^3}{3}}Z_{i,2}\right) +$$
$$+ \mathcal{L}_1\sigma\left(t_i, x\left(t_i\right)\right)\left(\frac{1}{2}Z_{i,1}^2\Delta_i - \frac{1}{2}\Delta_i\right). \tag{4.15}$$

For the Milstein scheme to work, we need up to second-order derivatives of μ and σ, which restricts the class of functions for which the method is viable. Furthermore, tedious calculations are often required to obtain these derivatives. It is possible to substitute them with numerical approximations, e.g., the finite difference method. However, doing so goes far beyond the scope of this dissertation, as we will only have to deal with the most common interest rate models, for which the derivatives exist and are quite nice and easy to compute.[18]

[17] By the Cholesky decomposition, $Z_{i,1} \coloneqq Y_{i,1}$ and $Z_{i,2} \coloneqq \sqrt{\frac{3}{4}}Y_{i,1} + \sqrt{\frac{1}{4}}Y_{i,2}$.

[18] For details on numerical approximations, see e.g. Kloeden & Platen (1992) or Glasserman (2004).

Chapter 5

Interest Rate Models

5.1 A Short History

Interest rate models are used to describe the evolution of interest rates and, consequently, the price of fixed income instruments such as bonds or swaps. An interest rate model picks an interest rate and makes assumptions about its dynamics. This is usually given by a differential equation. Interest rates can be assumed to be deterministic, as is the case in most equity models, however, theory and empirical evidence agree that a stochastic behaviour is more plausible. Recall from Section 2 that there are many interest rates to choose from, but once we make our choice, the rest is well-defined and the ordinary or stochastic differential equations can be transformed into each other.

Historically, the most popular choice for a state variable is the short rate $r(t)$ which is also the rate at which the money market instrument – the numéraire or bank account or bank deposit – grows. Although there exist models concentrating on the instantaneous forward rate, like the Heath et al. (1992), the majority of models are occupied with short rates.

Literature on short rate models is vast, and agents in the market often find themselves in a position where abundance makes it difficult to choose. When opting for a model, one must make mental notes about the properties and possible drawbacks of candidate models. Brigo & Mercurio (2006) list some of the most fundamental questions to consider when evaluating a model's merits:

1. **Is it an equilibrium or a no-arbitrage model?**

   A possible classification of interest rate models casts them into either equilibrium or no-arbitrage models. This is one of the first things to establish, and more often than

not the answer is a firm yes to no-arbitrage models, at least for pricing purposes. The main difference between the two lies in how each handles the yield curve currently observed on the market. While equilibrium models generally work with very few parameters and assume the yield curve to be *endogenous*, that is, an output of the model, no-arbitrage models do the exact opposite by taking the yield curve as an input and setting the parameters so that the whole yield curve is matched perfectly. It is evident that if we want our model to match market data, we have to opt for the latter. Nonetheless, for predictive purposes, equilibrium models produce more reliable results, due to their parsimonious nature. Moreover, for model development and model building, they prove to be a good starting point, with a few important features easy to comprehend and replicate. From the most often referenced models, the Vasicek (1977) and the Cox et al. (1985) are equilibrium, while the Hull & White (1990) and the Black & Karasinski (1991) are no-arbitrage models.

2. **What distribution does the short rate follow?**

   This is a fundamental question when creating or choosing a model, and has the highest impact on how rates or prices would evolve. One has to take into account that on the one hand, the best fit to empirical data is desirable, and on the other hand, some distributions are more appealing due to their advantageous properties. This question is thus closely linked to the following three:

3. **Can negative short rates occur?**

   Although recently the world took an interest in negative rates (Brigo & Mercurio (2001), Antonov et al. (2015)), for many decades it was highly undesirable for a model to produce negative short rates. Some model families were developed for the exact reason to preclude these. If, for example, the short rate follows a log-normal or a chi-squared distribution, like in the Black & Karasinski (1991) and Cox et al. (1985) models, the probability of getting negative rates is zero.

4. **Is there an analytical formula to compute the bond price?**

   One peculiarity of interest rate models is that we model the dynamics of the rate which in itself is not a traded instrument. The bond price, and specifically, the zero coupon bond price, however, is not only one of the most traded assets, but also serves as the discount factor – or the reciprocal of the numéraire. Hence it is paramount to be priced as precisely as possible. The surest way is an analytical formula which can be computed explicitly from present – and perhaps past – values

of the state variable. Unfortunately, there are not many models which guarantee the existence of such a formula, and it largely depends on the short rate's distribution.

5. **Is there an analytical formula to compute the bond option price?**

   Another problem is the existence of such a formula for bond options. There are even fewer models which can do this, and most of them are, unsurprisingly, equilibrium models with constant parameters. The Vasicek (1977), the Cox et al. (1985), and the no-arbitrage Hull & White (1990) are examples.

6. **What shapes of the yield curve can be reproduced?**

   When modelling interest rates, there is a natural desire to be able to replicate the most common shapes of historical or theoretical yield curves: the flat one, the increasing, the inverse and the humped. While most models perform well in this aspect, the Ho & Lee (1986), for example, is unable to cover all cases.

7. **Can the volatility surface observed on the market be reproduced?**

   Like when we wanted to match the yield curve by introducing no-arbitrage models, most equilibrium models have extensions to match the volatility term structure as well. Brigo & Mercurio (2006) warns, though, that this can be counter-effective for two reasons. First, some of the volatilities quoted on market are insignificant and incorporating them can result in overfitting, and second, the shape of the volatility term structure implied by such models is unrealistic.

8. **Can the model incorporate mean reversion?**

   One stylized fact of interest rates is that they tend to a long-term mean. It is called mean reversion and is present in many models, starting from the Vasicek (1977), through Cox et al. (1985) and Hull & White (1990) to Black & Karasinski (1991). The Ho & Lee (1986) model, on the other hand, is criticised for failing to include this characteristic.

9. **How easy is it to implement numerical schemes of the model?**

   Since very few models are analytically tractable, numerical methods are our best chance to solve them or price instruments in them. It is an important aspect, then, to consider how easily can such schemes, e.g. the Monte Carlo or tree methods, be adapted to the model.

Historically, the first model was proposed by Vasicek (1977) who assumed that the short rate follows an Ornstein–Uhlenbeck process:

$$dr(t) = \varkappa(\vartheta - r(t))\,dt + \sigma\,dW(t).$$

The Vasicek model is an equilibrium one and satisfies the mean reversion criterion. It is easy to see that the short rate is Gaussian[1], so it can take negative values. It was soon followed by Dothan (1978) who attempted to resolve some of these shortcomings. Their model is a simple geometric Brownian motion, i.e. log-normal, and there is no mean reversion, so even though it handled the negativity issue well, it proved to be a dead end for further improvements. The Ho & Lee (1986) model exhibited similar deficiencies, as though it was intended to fit the yield curve, it failed to resolve negative rates or include mean reversion.

One true breakthrough was the Cox et al. (1985) or CIR model that kept the mean reverting drift term, and remedied the negativity by introducing a $\sqrt{r(t)}$ factor in the diffusion term. Although it is still an equilibrium model, it paved the way for natural extensions to fit the term structure. Hull & White (1990) used a different approach and made the long-term mean ϑ time-dependent, meaning the model can fit the whole yield curve. It also kept the Gaussian property, which makes calculations easy but unfortunately does not preclude negative rates. Going one step further, by making σ and perhaps even $\varkappa$ time-dependent, we can fit the volatility term structure as well. This family of models is called the one-factor Gaussian short rate model, and has the nice property – as all previously mentioned models – of the existence of analytical formula for at least the zero coupon bond, but in the time-homogeneous cases, including Hull & White (1990), for the bond option as well.

Another family of models includes the Black et al. (1990), originally developed in discrete setting, and later generalised by Black & Karasinski (1991). The Black–Karasinski model is a log-normal version of the extended Vasicek, so it resolves negativity and is a no-arbitrage model, but unfortunately this comes with the disadvantage of not being analytically tractable which makes it impossible to obtain an explicit bond or bond option formula, meaning the Jamshidian decomposition does not work. Other non-linear extensions include quadratic, exponential, and quasi-Gaussian models.

Further research was done into developing extensions of existing models: the CIR for ex-

[1] See Section 5.3.

ample has many improvements including time-dependence (Maghsoodi (1996) or Jamshidian (1995)) or jumps and shifts (e.g. Brigo & Mercurio (2001) or Brigo & El-Bachir (2010)). Multi-factor extensions are also very common and exist for all one-factor models; see, e.g. Longstaff & Schwartz (1992) for a two-factor CIR, and Hull & White (1994*a*) for an application to Hull–White and other two-factor Markov models.

Another class of models is market models, the most famous being the LIBOR market model, which is beginning to lose its competitive edge with the recent LIBOR transition, but remains a good base for further research.

Stochastic volatility models comprise of at least two SDEs: one for the interest rate, and one for the diffusion term itself. Often they work with forward rates and forward volatilities instead of short ones. Perhaps the most widely used model today is the Stochastic Alpha-Beta-Rho, or SABR model (Hagan et al. (2002)). The most famous class of forward rate models is the Heath–Jarrow–Morton (HJM), proposed in the seminal work of Heath et al. (1992), which incorporates many short rate models. The Cheyette (2001) model is a Markov extension of the HJM and is also widely applied in industry.

Along the emergence of interest rate models, a lot of effort was put into finding ways to calculate bond prices in each of them. Jamshidian (1989) was the first to propose a formula for coupon-bearing bonds (and other sum-like pay-off function instruments, such as swaptions) if there exists an analytical one for zero coupon bonds. Wei (1997) introduced a duration-based simple approach for the Vasicek and CIR models, and Chen & Scott (1992) worked on a method for the two-factor CIR.

Similar enthusiasm surrounded option, and in particular, swaption pricing. Dybvig (1997) and Henrard (2003) both made significant contributions, while Heston (1993)'s paper on stochastic volatility models is one of the greatest breakthroughs in mathematical finance. As for swaption pricing, alongside the techniques detailed in Chapter 4, numerical approximations became popular. Collin-Dufresne & Goldstein (2002) and Schrager & Pelsser (2006) both worked on affine approximations, while Choi & Shin (2016) developed a technique called fast swaption pricing with the help of hyperplanes. Since path-dependent swaptions, and specifically, Bermudans, are really common on markets, the literature is vast on Bermudan swaption pricing, with remarkable contributions by Hagan (2000) or more recently, Gatarek & Jabłecki (2021).

Table 5.1, based on Brigo & Mercurio (2006) Table 3.1., concludes some of the best known short-rate models with some properties. V, H–L, D, CIR, HW, and BK denote the Vasicek, Ho–Lee, Dothan, CIR, Hull–White and Black–Karasinski models, respectively. E/N refers

to whether the model is equilibrium or no-arbitrage, the next column is whether negative rates are allowed, the next is distribution of the short rate where $\mathcal{N}$ is Gaussian, L$\mathcal{N}$ is log-normal, NCχ^2 is the non-central chi-squared distribution. AB refers to whether an analytical bond formula exists, AO is the same for an analytical option formula. Y and N denote "yes" and "no".

Model	Dynamics	E/N	$r > 0$	r dist.	AB	AO
V	$dr(t) = \varkappa(\vartheta - r(t))\,dt + \sigma\,dW(t)$	E	N	$\mathcal{N}$	Y	Y
H–L	$dr(t) = \vartheta(t)\,dt + \sigma\,dW(t)$	N	N	$\mathcal{N}$	Y	Y
D	$dr(t) = \alpha r(t)\,dt + \sigma r(t)\,dW(t)$	E	Y	L$\mathcal{N}$	Y	N
CIR	$dr(t) = \varkappa(\vartheta - r(t))\,dt + \sigma\sqrt{r(t)}\,dW(t)$	E	Y	NCχ^2	Y	Y
HW	$dr(t) = \varkappa(\vartheta(t) - r(t))\,dt + \sigma\,dW(t)$	N	N	$\mathcal{N}$	Y	Y
BK	$d\ln r(t) = (\vartheta(t) - \phi(t)\ln r(t))\,dt + \sigma\,dW(t)$	N	Y	L$\mathcal{N}$	N	N

Table 5.1: Short rate models (Brigo & Mercurio 2006, p. 57.).

It is apparent that among these models only the Vasicek, the CIR, and the Hull–White have an exact formula for both bond and bond option prices and at the same time exhibit mean reversion. Since we want to show the application of the Jamshidian decomposition to swaption pricing, in the rest of the dissertation we will only work with these three models, after making a short detour to stochastic differential equations.

5.2 Stochastic Differential Equations

To work with interest rate models, we might need to solve stochastic differential equations. We only work with one-factor Ito processes[2], so the general formula for an SDE is

$$dX(t) = \mu(t, X(t))\,dt + \sigma(t, X(t))\,dW(t)$$

where $W(t)$ is a Wiener process in its own filtration. Of course this differential form is a mere formalism and we *always* mean the implied integral form:

$$X(t) = X(0) + \int_0^t \mu(s, X(s))\,ds + \int_0^t \sigma(s, X(s))\,dW(s). \tag{5.1}$$

A solution of this SDE consists of a filtered probability space, a Wiener process W and

[2] X_t is an Ito process if it can be written in the integral form (5.1), where all terms exist and are well-defined.

an adapted process X, but henceforth when referring to a solution, we will only mean X. For a precise definition and further details, see Le Gall et al. (2016).

Generally, there is no guarantee that the SDE can be solved, but there are some regularity conditions for μ and σ to ensure existence.[3] Even if these assumptions hold, we are very rarely able to compute an explicit solution. Apart from a few special cases, the only easy-to-solve types are linear and Geometric Brownian Motion SDEs.

5.2.1 Linear SDEs

If the SDE is linear, we can easily compute a solution, as we will show. Consider the following SDE:

$$dX(t) = AX(t) + B(t)\,dt + C(t)\,dW(t). \tag{5.2}$$

Proposition. 5.2.1. *The SDE in (5.2) has a unique solution*[4]

$$X(t) = e^{At}X(0) + \int_0^t e^{A(t-s)}B(s)\,ds + \int_0^t e^{A(t-s)}C(s)\,dW(s). \tag{5.3}$$

Proof. The uniqueness follows from the Lipschitz regularity condition, which the linear SDE satisfies.

To prove the existence, apply Ito's lemma to $f(t, X) \coloneqq e^{-At}X$:

$$\begin{aligned} df &= -Ae^{-At}X(t)\,dt + e^{-At}\,dX(t) = \\ &= e^{-At}\left(-AX(t) + AX(t) + B(t)\right)dt + e^{-At}C(t)\,dW(t). \end{aligned}$$

Writing in integral form and then multiplying by e^{At}:

$$e^{-At}X(t) - e^{-A0}X(0) = \int_0^t e^{-As}B(s)\,ds + \int_0^t e^{-As}C(s)\,dW(s)$$

$$X(t) = e^{At}X(0) + \int_0^t e^{A(t-s)}B(s)\,ds + \int_0^t e^{A(t-s)}C(s)\,dW(s).$$

□

[3] E.g. a Lipschitz and a linear growth condition (Le Gall et al. 2016, pp. 212-213.).

[4] The formula can be extended to time-dependent $A(t)$, see e.g. Karatzas & Shreve (2012) pp. 360-361.

Linear SDEs of this form are sometimes referred to as Ornstein–Uhlenbeck processes, and are very common among short rate models.

5.2.2 Geometric Brownian Motion

The other explicitly solvable SDE type is the so-called Geometric Brownian Motion (GBM). Consider the SDE:

$$dX(t) = \mu(t) X(t) + dt + \sigma(t) X(t) dW(t). \tag{5.4}$$

Proposition. 5.2.2. *The SDE in (5.4) has a unique solution*

$$X(t) = X(0) \exp\left\{ \int_0^t \mu(s) - \frac{1}{2}\sigma(s)^2 ds + \int_0^t \sigma(s) dW(s) \right\}. \tag{5.5}$$

Proof. The uniqueness again follows from the Lipschitz regularity condition.

To prove the existence, apply Ito's lemma to $f(t, X) \coloneqq \ln X$.

$$df = \frac{1}{X(t)} dX(t) - \frac{1}{2X(t)^2} d[X(t)] = \mu(t) dt + \sigma(t) dW(t) - \frac{1}{2}\sigma(t)^2 dt =$$
$$= \mu(t) - \frac{1}{2}\sigma(t)^2 dt + \sigma(t) dW(t).$$

Writing in integral form and then taking exponential:

$$\ln X(t) - \ln X(0) = \int_0^t \mu(s) - \frac{1}{2}\sigma(s)^2 ds + \int_0^t \sigma(s) dW(s)$$
$$X(t) = X(0) \exp\left\{ \int_0^t \mu(s) - \frac{1}{2}\sigma(s)^2 ds + \int_0^t \sigma(s) dW(s) \right\}$$

□

Remark. Under certain regularity conditions[5], it is clear that the stochastic integral is Gaussian with mean 0 and variance $\int_0^t \sigma(s)^2 ds$. Thus $X(t)$ is log-normal with mean and

[5] If e.g. σ is square-integrable and adapted.

variance[6]

$$\mathbb{E}\left[X\left(t\right)\right] = X\left(0\right)\exp\left\{\int_0^t \mu\left(s\right)ds\right\} \tag{5.6}$$

$$\mathbb{D}^2\left[X\left(t\right)\right] = \mathbb{E}\left[X\left(t\right)\right]^2\left(\exp\left\{\int_0^t \sigma\left(s\right)^2 ds\right\} - 1\right). \tag{5.7}$$

GBMs are very popular in financial modelling, especially in equity models, but when the short rate is Gaussian, e.g. in the Vasicek or Hull–White models, then zero coupon bonds are log-normal and thus their dynamics can be written in the form of (5.4).

5.3 The Vasicek Model

When one is working with market data, the best practice is to use short rates as state variables. It comes with the advantage of having to deal with only one time variable. This is one reason why the majority of interest rates models centre around the dynamics of short rates, even if most of these models are incorporated in the (forward rates) Heath–Jarrow–Morton family. In the next few sections, we will present three of the most famous one-factor short-rate models: the Vasicek, the Hull–White, and the Cox–Ingersoll–Ross models.

The Vasicek (1977) model is one of the earliest interest rate models and is an equilibrium model, meaning that it can be derived from some assumptions, and it *cannot* fit the current yield curve perfectly. To remedy this, numerous attempts have been made to switch to an exogenous rather than an endogenous yield curve, the most famous being the Hull & White (1990) extended Vasicek model, which we will examine later.

In the Vasicek model, the short rate follows a mean reverting Ornstein–Uhlenbeck process to capture the empirical fact that interest rates tend to a long-term equilibrium level:

$$dr\left(t\right) = \varkappa\left(\vartheta - r\left(t\right)\right)dt + \sigma\, dW\left(t\right). \tag{5.8}$$

It is evident that the SDE is linear, with a flat short volatility σ, long-term mean ϑ, and mean reversion speed $\varkappa$. It is also clear that although the short rate will be centred

[6] If Y is Gaussian with mean μ and variance σ^2 then $X \coloneqq e^Y$ is log-normal with mean $\mathbb{E}\left[X\right] = \exp\left\{\mu + \dfrac{\sigma^2}{2}\right\}$ and variance $\mathbb{D}^2\left[X\right] = \mathbb{E}\left[X\right]^2\left(\exp\left\{\sigma^2\right\} - 1\right)$. See Medvegyev (2002) pp. 285-286.

around its – most likely positive – mean, the formulation of the SDE does not exclude the possibility of negative interest rates, a handicap that later models, for example, the Cox et al. (1985) tried to correct.

Since the SDE is linear, we can apply the formula in (5.3) to solve it. With $A \coloneqq -\varkappa$, $B(t) \coloneqq \varkappa\vartheta$ and $C(t) \coloneqq \sigma$, the solution for $r(t)$ is:

$$\begin{aligned} r(t) &= e^{-\varkappa t} r(0) + \int_0^t e^{-\varkappa(t-s)} \varkappa\vartheta \, ds + \int_0^t e^{-\varkappa(t-s)} \sigma \, dW(s) = \\ &= e^{-\varkappa t} r(0) + \frac{\varkappa\vartheta}{\varkappa} \left[e^{-\varkappa(t-s)} \right]_0^t + \int_0^t e^{-\varkappa(t-s)} \sigma \, dW(s) = \\ &= e^{-\varkappa t} r(0) + \left(1 - e^{-\varkappa t}\right) \vartheta + \int_0^t e^{-\varkappa(t-s)} \sigma \, dW(s) . \end{aligned} \tag{5.9}$$

Since the integrand is bounded and adapted, the short rate is Gaussian with mean $e^{-\varkappa t} r(0) + \left(1 - e^{-\varkappa t}\right) \vartheta$. Note that this mean is a convex combination of the initial short rate and the long-term mean, and as $t \to \infty$, it tends to ϑ. Hence the Vasicek indeed is a mean reverting process. To compute the variance, we use the Ito-isometry:

$$\mathbb{D}^2 [r(t)] = \int_0^t e^{-2\varkappa(t-s)} \sigma^2 \, ds = \frac{\sigma^2}{2\varkappa} \left(1 - e^{-2\varkappa t}\right) .$$

It increases with time and tends to $\dfrac{\sigma^2}{2\varkappa}$.

Similarly, we can write $r(T)$ as a function of $r(t)$ where $t \leq T$:

$$r(T) = e^{-\varkappa(T-t)} r(t) + \left(1 - e^{-\varkappa(T-t)}\right) \vartheta + \int_t^T e^{-\varkappa(T-s)} \sigma \, dW(s) . \tag{5.10}$$

Of course it remains Gaussian, and its conditional mean and variance are:

$$\begin{aligned} \mathbb{E}[r(T) | \mathcal{F}_t] &= e^{-\varkappa(T-t)} r(t) + \left(1 - e^{-\varkappa(T-t)}\right) \vartheta \\ \mathbb{D}^2 [r(T) | \mathcal{F}_t] &= \frac{\sigma^2}{2\varkappa} \left(1 - e^{-2\varkappa(T-t)}\right) . \end{aligned}$$

5.3.1 Zero Coupon Bond in the Vasicek Model

To price swaptions, we need formulae for zero coupon bonds rather than interest rate levels. Numerous methods are known to derive the discount bond formula, Mamon (2004) gives a comprehensive summary of three of these: derivation from the short rate distribution, derivation from the partial differential equation (PDE) of the bond itself, known as the Term Structure Equation (TSE), and finally, derivation from the HJM framework. We will apply the TSE method to the Cox–Ingersoll–Ross model and the HJM method to the Hull–White model. Therefore, though Vasicek (1977) uses the PDE method, we, in accordance with Andersen & Piterbarg (2010), will present the first one: derivation from the short rate distribution.

It is known from earlier that the price of the zero coupon bond is given by

$$P(t,T) = \mathbb{E}\left[\exp\left\{-\int_t^T r(s)\,ds\right\}|\mathcal{F}_t\right].$$

Inserting (5.10) into the integral, and in the first step, using the stochastic Fubini theorem[7] since the integrand is bounded and predictable:

$$\begin{aligned}\int_t^T r(s)\,ds &= \vartheta(T-t) + (r(t)-\vartheta)\int_t^T e^{-\varkappa(T-s)}\,ds + \sigma\int_t^T\int_t^s e^{-\varkappa(s-u)}\,dW(u)\,ds = \\ &= \vartheta(T-t) + \frac{r(t)-\vartheta}{\varkappa}\left(1-e^{-\varkappa(T-t)}\right) + \sigma\int_t^T\int_u^T e^{-\varkappa(s-u)}\,ds\,dW(u) = \\ &= \vartheta(T-t) + \frac{r(t)-\vartheta}{\varkappa}\left(1-e^{-\varkappa(T-t)}\right) + \frac{\sigma}{\varkappa}\int_t^T 1-e^{-\varkappa(T-u)}\,dW(u).\end{aligned}$$

Since the integrand is again adapted and bounded, the expression is Gaussian with conditional mean

$$\mathbb{E}\left[\int_t^T r(s)\,ds\,|\mathcal{F}_t\right] = \vartheta(T-t) + \frac{r(t)-\vartheta}{\varkappa}\left(1-e^{-\varkappa(T-t)}\right)$$

[7] See Medvegyev (2014) Theorem 7.9.

and variance

$$\mathbb{D}^2\left[\int_t^T r(s)\,ds\,|\mathcal{F}_t\right] = \frac{\sigma^2}{\varkappa^2}\int_t^T \left(1-e^{-\varkappa(T-u)}\right)^2 du = \frac{\sigma^2}{\varkappa^2}\left[u-2e^{-\varkappa(T-u)}+e^{-2\varkappa(T-u)}\right]_t^T =$$
$$= \frac{\sigma^2}{\varkappa^2}\left(T-t-\frac{2}{\varkappa}\left(1-e^{-\varkappa(T-t)}\right)+\frac{1}{2\varkappa}\left(1-e^{-2\varkappa(T-u)}\right)\right).$$

Now it is clear that the exponential follows a log-normal distribution and from the expected value formula for log-normal variables, the zero coupon bond price is:

$$P(t,T) = \mathbb{E}\left[\exp\left\{-\int_t^T r(s)\,ds\right\}|\mathcal{F}_t\right] =$$
$$= \exp\left\{-\mathbb{E}\left[\int_t^T r(s)\,ds\,|\mathcal{F}_t\right]+\frac{1}{2}\mathbb{D}^2\left[\int_t^T r(s)\,ds\,|\mathcal{F}_t\right]\right\} = \exp\left\{-r(t)\,\frac{1-e^{-\varkappa(T-t)}}{\varkappa}\right\}\times$$
$$\times\exp\left\{-(T-t)\left(\vartheta-\frac{\sigma^2}{2\varkappa^2}\right)+\left(\vartheta-\frac{\sigma^2}{2\varkappa^2}\right)\frac{1-e^{-\varkappa(T-t)}}{\varkappa}-\frac{\sigma^2}{4\varkappa}\left(\frac{1-e^{-\varkappa(T-t)}}{\varkappa}\right)^2\right\} =$$
$$=: \exp\left\{-r(t)\,B(t,T)+A(t,T)\right\}$$

where

$$B(t,T) := \frac{1-e^{-\varkappa(T-t)}}{\varkappa}$$
$$A(t,T) := \left(\vartheta-\frac{\sigma^2}{2\varkappa^2}\right)(B(t,T)-(T-t))-\frac{\sigma^2}{4\varkappa}B(t,T)^2. \tag{5.11}$$

So the logarithm of the zero coupon bond is an affine function of the short rate. This type of model is thus called an affine model.

5.3.2 Zero Coupon Bond Option in the Vasicek Model

We just showed that there exists a variable $r(t)$ in which the zero coupon bond $P(t,T)$ is monotone; moreover, we know its dynamics. The assumptions of the Jamshidian trick thus hold, so we can decompose the swaption into a basket of zero coupon bond options. Now it is our objective to price these options. Fortunately, in the Vasicek model, there exist analytical formulae for the call and put options, very similar to that of the Black–Scholes world. Later, in the general Gaussian model, we will elaborate more on this finding; here, we only present the formula for the Vasicek case without proof. For further

details, see e.g. Brigo & Mercurio (2006).

Proposition. 5.3.1 (Black formula for Vasicek model)**.** *Let* $V_{call}(T) \coloneqq (P(T,T^*) - K)^+$ *denote the pay-off function at expiry date* T *of the zero coupon bond call option, with maturity* T^* *and strike* K*.*

In the Vasicek model, the option's value at time $t \leq T$ *is:*

$$\begin{aligned} \boldsymbol{ZBC}(t) &= P(t,T^*)\,\Phi(d_+) - P(t,T)\,K\Phi(d_-) \\ d_\pm &= \frac{\ln\left(P(t,T^*)/(KP(t,T))\right) \pm \nu/2}{\sqrt{\nu}} \\ \nu &= \sigma^2 \frac{1-e^{-2\varkappa(T-t)}}{2\varkappa} B(T,T^*)^2 . \end{aligned} \tag{5.12}$$

Remark. Similarly, for the zero coupon bond put option:

$$\mathbf{ZBP}(t) = P(t,T)\,K\Phi(-d_-) - P(t,T^*)\,\Phi(-d_+) . \tag{5.13}$$

Now we have every tool to find an analytical price for the swaption with the Jamshidian decomposition.

5.3.3 Binomial Tree for the Vasicek Model

This section presents a binomial tree approach applied to the Vasicek model and is based on Nelson & Ramaswamy (1990).

Consider a simple binomial tree with a single node as the origin and two children nodes, as in Figure 5.1. In the origin, the short rate takes the value r, and the tree being an additive version of the CRR model, let the up and down step from each node be $\sigma\sqrt{\Delta}$ and $-\sigma\sqrt{\Delta}$, respectively, where Δ denotes the equidistant time step on the grid. By this construction, the tree is also equidistant vertically, and the vertical step is constant over time. Let $p(r)$ denote the probability of moving upward from a node with value r. Note that this depends on the value taken, rather than the position in the grid. Let

$$p(r) \coloneqq \frac{1}{2} + \frac{\varkappa(\vartheta - r)\sqrt{\Delta}}{2\sigma}.$$

Note that in certain cases, where the time step is too large or the interest rate is far from its long-term mean, this probability can take values outside of the $[0,1]$ interval. To remedy this, we can either cut the tree where this would take place, or use a floor and

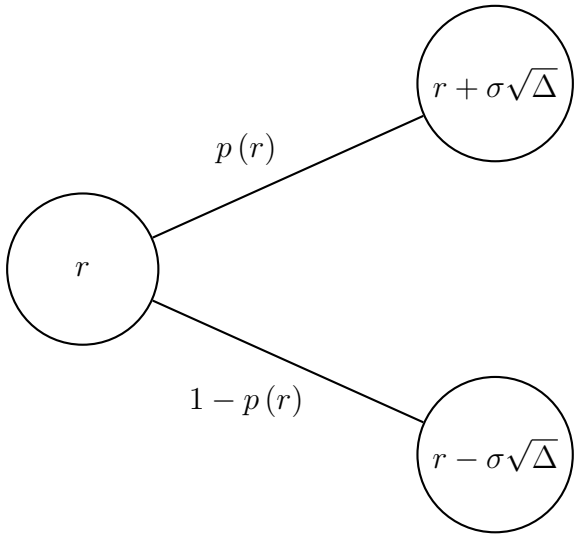


Figure 5.1: Building block of binomial tree.

ceiling for the probability, or even leave it this way, as it is known that convergence is guaranteed, regardless.

The first two moments of the change in the interest rate over the Δ time step are

$$\mathbb{E}[\Delta r] = \left(\frac{1}{2} + \frac{\varkappa(\vartheta - r)\sqrt{\Delta}}{2\sigma}\right)\sigma\sqrt{\Delta} - \left(\frac{1}{2} - \frac{\varkappa(\vartheta - r)\sqrt{\Delta}}{2\sigma}\right)\sigma\sqrt{\Delta} = \varkappa(\vartheta - r)\Delta,$$
$$\mathbb{E}\left[(\Delta r)^2\right] = p(r)\left(\sigma\sqrt{\Delta}\right)^2 + (1 - p(r))\left(-\sigma\sqrt{\Delta}\right)^2 = \left(\sigma\sqrt{\Delta}\right)^2 = \sigma^2\Delta.$$

If we impose restrictions on the probability, the second moment does not change, and the first becomes $-\sigma\sqrt{\Delta}$ on the upper end of the tree and its negative on the lower end.

Recall that these results are perfectly in line with what we know from the SDE of the model. We have thus successfully replicated a discretised version of the Vasicek model on a lattice tree.

To price a swaption, we first create an equidistant grid spanning from today (time 0) to the last payment date (T_n) of the swaption, so that T_i and T_{i+1} are consecutive dates on the grid, including the expiration date T_0 and every payment date.[8] Second, we construct the tree of rates following the method above. To price swaps, we need a bunch of zero coupon bonds, each starting at T_0 and ending at one of the payment dates, i.e. the nodes

[8]This essentially means that we can only price swaptions where payment dates are equidistant, but this already covers most cases and is not hard to generalise.

on the tree. Recall that the formula for pricing zero coupon bonds generally is

$$P(t,T) = \mathbb{E}\left[\exp\left\{-\int_t^T r(s)\,ds\right\}|\mathcal{F}_t\right].$$

The expectation at a node comes from taking the expectation of the values taken at the children nodes, with the up and down probabilities. On each time step of the grid, we assume the short rate to be flat and left-continuous. As a consequence, the discount factor – i.e. the zero coupon bond – between two consecutive nodes becomes

$$\begin{aligned} P(T_i,T_{i+1})_j &= \mathbb{E}\left[\exp\left\{-r_{j,i}\Delta\right\}|\mathcal{F}_{T_i}\right] = \exp\left\{-r_{j,i}\Delta\right\}\left(p\left(r_{j,i}\right)+\left(1-p\left(r_{j,i}\right)\right)\right) = \\ &= \exp\left\{-r_{j,i}\Delta\right\}. \end{aligned}$$

where j is the vertical index and refers to the state in which the variable is, and i is the time index, while $r_{j,i}$ denotes the value of the short rate at time T_i taken in state j. By induction, the discount factor between node $T_{j,i}$ and T_k with $i < k$ is

$$P(T_i,T_k)_j = \exp\left\{-r_{j,i}\Delta\right\}\mathbb{E}\left[P\left(T_{i+1},T_k\right)\right] = P(T_i,T_{i+1})_j\,\mathbb{E}\left[P\left(T_{i+1},T_k\right)\right]. \tag{5.14}$$

To conclude, we have to create trees for $i = 1,\ldots,n$ period zero coupon bonds by the above recursion, use these at each node at time T_0 to price the swap with (3.1), and then evaluate the swaption. Then from these nodes, using the calculated zero coupon bond prices, roll back to the origin to get the swaption's present value.

5.3.4 Monte Carlo Methods in the Vasicek Model

To run a Monte Carlo simulation, we first have to generate trajectories. All techniques listed in Section 4.3 work in the Vasicek model, and we detail the explicit Euler–Maruyama, the Euler scheme with linear drift and the Milstein second-order scheme here. Though in theory, the Milstein scheme is bound to converge faster, Andersen & Piterbarg (2010) insists that a second-order scheme is highly overachieving in this very simple setting of a linear model. We have to agree with the authors in that the Milstein scheme does not add much to the first-order discretisation, indeed.

The Explicit Euler–Maruyama Scheme

The simplest discretisation is, as can be deduced from an earlier chapter:

$$r_{i+1} = r_i + \varkappa\left(\vartheta - r_i\right)\Delta_i + \sigma\sqrt{\Delta_i}Z_i \tag{5.15}$$

where Z_i is a standard Gaussian random variable.

Euler Scheme with Linear Drift

Consider the transformation $Y(t) \coloneqq e^{\varkappa t}r(t) - \varkappa\int_0^t \vartheta e^{\varkappa u}\,du$. From Ito's lemma:

$$\begin{aligned} dY(t) &= \left(\varkappa r(t)\,e^{\varkappa t} - \varkappa\vartheta e^{\varkappa t}\right)dt + e^{\varkappa t}\,dr(t) = \\ &= \left(\varkappa r(t)\,e^{\varkappa t} - \varkappa\vartheta e^{\varkappa t} + e^{\varkappa t}\varkappa\vartheta - e^{\varkappa t}\varkappa r(t)\right)dt + e^{\varkappa t}\sigma\,dW(t) = \\ &= e^{\varkappa t}\sigma\,dW(t). \end{aligned}$$

Apply an explicit Euler–Maruyama scheme to Y:

$$Y_{i+1} = Y_i + e^{\varkappa t_i}\sigma\sqrt{\Delta_i}Z_i$$

where Z_i is a Gaussian variable with mean 0 and variance 1.

Now express $r(t)$ and $r(t+\Delta_i)$ as a function of $Y(t)$ and $Y(t+\Delta_i)$:

$$\begin{aligned} r_i &= e^{-\varkappa t_i}\left(Y_i + \varkappa\int_0^{t_i}\vartheta e^{\varkappa u}\,du\right), \\ r_{i+1} &= e^{-\varkappa(t_i+\Delta_i)}\left(Y_{i+1} + \varkappa\int_0^{t_i+\Delta_i}\vartheta e^{\varkappa u}\,du\right) = \\ &= e^{-\varkappa\Delta_i}e^{-\varkappa t_i}\left(Y_i + \varkappa\vartheta\int_0^{t_i}e^{\varkappa u}\,du\right) + e^{-\varkappa\Delta_i}e^{-\varkappa t_i}e^{\varkappa t_i}\sigma\sqrt{\Delta_i}Z_i + e^{-\varkappa\Delta_i}e^{-\varkappa t_i}\varkappa\vartheta\int_{t_i}^{t_i+\Delta_i}e^{\varkappa u}\,du = \\ &= e^{-\varkappa\Delta_i}r_i + e^{-\varkappa\Delta_i}\sigma\sqrt{\Delta_i}Z_i + \frac{\varkappa\vartheta}{\varkappa}e^{-\varkappa(t_i+\Delta_i)}\left[e^{\varkappa u}\right]_{t_i}^{t_i+\Delta_i} = \\ &= e^{-\varkappa\Delta_i}r_i + e^{-\varkappa\Delta_i}\sigma\sqrt{\Delta_i}Z_i + \left(1 - e^{-\varkappa\Delta_i}\right)\theta = \\ &= e^{-\varkappa\Delta_i}r_i + \left(1 - e^{-\varkappa\Delta_i}\right)\theta + e^{-\varkappa\Delta_i}\sigma\sqrt{\Delta_i}Z_i. \end{aligned} \tag{5.16}$$

This is the modified Euler discretisation for the Vasicek model. Note that it is exactly the

same as if we used the explicit Euler–Maruyama method for the solution of the Vasicek SDE, conditional on the "previous" filtration, i.e. (5.10).

The Milstein Scheme

First, we establish $\mu(t, r(t)) \coloneqq \varkappa\vartheta - \varkappa r(t)$ and $\sigma(t, r(t)) \coloneqq \sigma$. Now we have to apply the $\mathcal{L}_0$ and $\mathcal{L}_1$ operators to each of them:

$$\begin{aligned}
\mathcal{L}_0\mu &= \left(\frac{\partial}{\partial t} + \mu(t,r)\frac{\partial}{\partial r} + \frac{1}{2}\sigma(t,r)^2\frac{\partial^2}{\partial r^2}\right)\mu(t,r) = 0 - \mu(t,r)\varkappa + 0 = -\varkappa^2(\vartheta - r(t)) \\
\mathcal{L}_1\mu &= \sigma(t,r)\frac{\partial}{\partial r}\mu(t,r) = -\varkappa\sigma \\
\mathcal{L}_0\sigma &= \left(\frac{\partial}{\partial t} + \mu(t,r)\frac{\partial}{\partial r} + \frac{1}{2}\sigma(t,r)^2\frac{\partial^2}{\partial r^2}\right)\sigma(t,r) = 0 \\
\mathcal{L}_1\sigma &= \sigma(t,r)\frac{\partial}{\partial r}\sigma(t,r) = 0.
\end{aligned}$$

Inserting these into (4.15), we get

$$r_{i+1} = r_i + \varkappa(\vartheta - r_i)\Delta_i + \sigma\sqrt{\Delta_i}Z_{i,1} - \frac{\Delta_i^2}{2}\varkappa^2(\vartheta - r_i) - \varkappa\sigma\sqrt{\frac{\Delta_i^3}{3}}Z_{i,2} \tag{5.17}$$

where $Z_{i,1}$ and $Z_{i,2}$ are standard Gaussian variables with correlation $\frac{3}{4}$.

5.4 The One-Factor Gaussian Short Rate Model

The Gaussian model is a generalisation of the Vasicek model: it is a mean reverting short-rate process, with time-varying coefficients and short volatility independent of the short rate. It takes the form

$$dr(t) = \varkappa(t)(\vartheta(t) - r(t))\,dt + \sigma_r(t)\,dW(t). \tag{5.18}$$

Although at first glance all $\varkappa(t), \vartheta(t)$ and $\sigma_r(t)$ could be exogenous, we desire that the short rate be consistent with the initial spot curve and to ensure that we have to make some assumptions. It is shown (e.g. Andersen & Piterbarg (2010)) that in this case the SDE corresponds to a Heath–Jarrow–Morton model.

$$df(t,T) = \sigma_f(t,T)\int_t^T \sigma_f(t,u)\,du\,dt + \sigma_f(t,T)\,dW(t) \tag{5.19}$$

$$\sigma_f(t,T) = \sigma_r(t) \exp\left\{ -\int_t^T \varkappa(u)\, du \right\}$$

and that the long-term mean must satisfy the following:

$$\vartheta(t) = \frac{1}{\varkappa(t)} \frac{\partial f(0,t)}{\partial t} + f(0,t) + \frac{1}{\varkappa(t)} \int_0^t \exp\left\{ -2\int_u^t \varkappa(s)\, ds \right\} \sigma_r(u)^2\, du\,.$$

5.4.1 Zero Coupon Bond in the Gaussian Model

We wish to eliminate the forward rate from the SDE, so we change to a new variable $x(t) := r(t) - f(0,t)$. To derive the SDE for $x(t)$ and a zero coupon bond formula, we follow Andersen & Piterbarg (2010)[9].

Proposition. 5.4.1. *The process $x(t)$ follows the SDE:*

$$\left.\begin{aligned} dx(t) &= (y(t) - \varkappa(t)\, x(t))\, dt + \sigma_r(t)\, dW(t) \\ x(0) &= 0 \end{aligned}\right\} \tag{5.20}$$

where $y(t) = \int_0^t \exp\left\{ -2\int_u^t \varkappa(s)\, ds \right\} \sigma_r(u)^2\, du.$

The zero coupon bond formula is

$$\begin{aligned} P(t,T) &= \frac{P(0,T)}{P(0,t)} \exp\left\{ -x(t)\, G(t,T) - \frac{1}{2} y(t)\, G(t,T)^2 \right\} \\ G(t,T) &= \int_t^T \exp\left\{ -\int_t^u \varkappa(s)\, ds \right\} du\,. \end{aligned} \tag{5.21}$$

Proof. Let $K(t) := \int_0^t \varkappa(u)\, du$, $g(t) := \sigma_r(t)\, e^{K(t)}$, and $h(t) := e^{-K(t)}$. In this way, the forward volatility is separable: $\sigma_f(t,T) = g(t)\, h(T)$. Integrate (5.19) from 0 to t:

$$f(t,T) - f(0,T) = \int_0^t \sigma_f(u,T) \int_u^T \sigma_f(u,s)\, ds\, du + \int_0^t \sigma_f(u,T)\, dW(u) =$$

[9] Proposition 10.1.7.

$$= \int_0^t g(u) h(T) \int_u^T g(u) h(s) \, ds \, du + \int_0^t g(u) h(T) \, dW(u) =$$

$$= h(T) \int_0^t g(u)^2 \int_u^T h(s) \, ds \, du + h(T) \int_0^t g(u) \, dW(u) . \tag{5.22}$$

Let

$$x(t) \coloneqq r(t) - f(0,t) = f(t,t) - f(0,t) = h(t) \int_0^t g(u)^2 \int_u^t h(s) \, ds \, du + h(t) \int_0^t g(u) \, dW(u) .$$

By the Leibniz integral rule[10]

$$\begin{aligned} dx(t) = & \left(h'(t) \int_0^t g(u)^2 \int_u^t h(s) \, ds \, du + h(t)^2 \int_0^t g(u)^2 \, du \right) dt + \\ & + h'(t) \int_0^t g(u) \, dW(u) \, dt + h(t) g(t) \, dW(t) = \\ & = \left(\frac{h'(t)}{h(t)} x(t) + y(t) \right) dt + h(t) g(t) \, dW(t) = \\ & = (y(t) - \varkappa(t) x(t)) \, dt + \sigma_r(t) \, dW(t) . \end{aligned}$$

We used the fact that

$$h(t)^2 \int_0^t g(u)^2 \, du = \exp\left\{ -2 \int_0^t \varkappa(u) \, du \right\} \int_0^t \sigma_r(u)^2 \exp\left\{ 2 \int_0^u \varkappa(s) \, ds \right\} du =$$

$$= \int_0^t \sigma_r(u)^2 \exp\left\{ -2 \int_u^t \varkappa(s) \, ds \right\} du =: y(t) .$$

From (5.22)

$$f(t,T) = f(0,T) + \frac{h(T)}{h(t)} x(t) + h(T) \int_0^t g(u)^2 \int_u^T h(s) \, ds \, du - h(T) \int_0^t g(u)^2 \int_u^t h(s) \, ds \, du =$$

[10] $\dfrac{\partial}{\partial x} \displaystyle\int_{a(x)}^{b(x)} f(x,t) \, dt = f(x, b(x)) \dfrac{d}{dx} b(x) - f(x, a(x)) \dfrac{d}{dx} a(x) + \displaystyle\int_{a(x)}^{b(x)} \dfrac{\partial}{\partial x} f(x,t) \, dt .$

$$= f(0,T) + \frac{h(T)}{h(t)} x(t) + h(T) \int_0^t g(u)^2 \, du \int_t^T h(s) \, ds =$$

$$= f(0,T) + \frac{h(T)}{h(t)} \left(x(t) + \frac{y(t)}{h(t)} \int_t^T h(s) \, ds \right). \tag{5.23}$$

The bond formula, as shown in Chapter 2, is

$$P(t,T) = \exp\left\{ - \int_t^T f(t,u) \, du \right\}.$$

Inserting in the formula in (5.23) and computing the components:

$$\exp\left\{ - \int_t^T f(0,u) \, du \right\} = \exp\left\{ - \int_0^T f(0,u) \, du + \int_0^t f(0,u) \, du \right\} =$$

$$= \exp\left\{ - \int_0^T f(0,u) \, du \right\} \cdot \exp\left\{ - \left(- \int_0^t f(0,u) \, du \right) \right\} =$$

$$= \frac{P(0,T)}{P(0,t)}.$$

$$\int_t^T \frac{h(u)}{h(t)} x(t) \, du = x(t) \int_t^T \exp\left\{ - \int_t^u \varkappa(s) \, ds \right\} du =: x(t) \, G(t,T).$$

$$\int_t^T \frac{h(u)}{h(t)^2} y(t) \int_t^u h(s) \, ds \, du = y(t) \int_t^T \frac{h(u)}{h(t)} \int_t^u \frac{h(s)}{h(t)} \, ds \, du =$$

$$= \frac{1}{2} y(t) \int_t^T \frac{h(u)}{h(t)} \, du \int_t^T \frac{h(s)}{h(t)} \, ds =: \frac{1}{2} y(t) \, G(t,T)^2$$

where the $\frac{1}{2}$ comes from integration over a triangle.

So the bond reconstitution formula is as follows:

$$P(t,T) = \frac{P(0,T)}{P(0,t)} \exp\left\{-x(t)\,G(t,T) - \frac{1}{2}y(t)\,G(t,T)^2\right\}.$$

□

Now we have a formula for pricing zero coupon bonds in the future. Note that these are, of course, random variables, depending on the path – or more precisely, only the time t value – of the driving process $x(t)$. The fact that it only depends on the value taken at t suggests that the process for the zero coupon bond is Markovian[11].

Remark. The SDE for the zero coupon bond in the Gaussian model is:

$$dP(t,T) = P(t,T)\,r(t)\,dt - P(t,T)\,\sigma_P(t,T)\,dW(t) \tag{5.24}$$

where $\sigma_P(t,T) = \sigma_r(t)\,G(t,T)$.

Proof. The Gaussian model is known to come from a Markovian Heath–Jarrow–Morton model, where the following holds for the zero coupon bond dynamics:

$$dP(t,T) = P(t,T)\,r(t)\,dt - P(t,T)\,\sigma_P(t,T)\,dW(t)$$

where $\sigma_P(t,T) = \int_t^T \sigma_f(t,u)\,du$. We also know that $\sigma_f(t,u) = g(t)\,h(u)$ is separable, with g and h defined above. Using this, we get:

$$\sigma_P(t,T) = \int_t^T g(t)\,h(u)\,du = \sigma_r(t) \int_t^T e^{-\int_t^u \varkappa(s)\,ds}\,du = \sigma_r(t)\,G(t,T).$$

That concludes the proof. □

5.4.2 Zero Coupon Bond Option in the Gaussian Model

Note that $P(t,T)$ is a GBM, so it must be log-normal[12]. In this case, similarly as in the famous Black–Scholes–Merton model for equity options, there exists an analytical formula, called the Black formula, to price options on zero coupon bonds.

[11] Indeed, it follows from the separable HJM. See Andersen & Piterbarg (2010).
[12] See Section 5.5.

Proposition. 5.4.2 (Black formula). *Let* $V_{call}(T) \coloneqq (P(T,T^*) - K)^+$ *denote the pay-off function at expiry date* T *of the zero coupon bond call option, with maturity* T^* *and strike* K*.*

In the Gaussian model, the option value at time $t \leq T$ *is:*

$$\begin{aligned} \boldsymbol{ZBC}(t) &= P(t,T^*)\,\Phi(d_+) - P(t,T)\,K\Phi(d_-) \\ d_\pm &= \frac{\ln\left(P(t,T^*)/(KP(t,T))\right) \pm \nu/2}{\sqrt{\nu}} \\ \nu &= \int_t^T |\sigma_P(u,T^*) - \sigma_P(u,T)|^2\,du\,. \end{aligned} \tag{5.25}$$

Note that it is exactly the same as the original BSM, with $S(t) \coloneqq P(t,T^*)$. It is also apparent that its SDE must not contain a drift term. In fact, the proof uses a change of measure to the T-forward measure under which $P(t,T,T^*)$ is a martingale, while its variance is given by ν. The complete proof can be found in Andersen & Piterbarg (2010) and Black & Scholes (1973).

Remark. The put option price can be derived similarly and is given by

$$\mathbf{ZBP}(t) = P(t,T)\,K\Phi(-d_-) - P(t,T^*)\,\Phi(-d_+)\,. \tag{5.26}$$

5.5 The Hull–White Model

The simplest version of the Gaussian model operates with flat mean reversion and volatility parameter, but keeps the mean as a function of time. It was developed by Hull & White (1990) and is thus often referred to as the Hull–White model or even the extended Vasicek model since it is a no-arbitrage extension of the latter. The SDE in (5.18) is rewritten as

$$dr(t) = \varkappa\left(\vartheta(t) - r(t)\right)dt + \sigma_r\,dW(t)$$

and the long-term mean is

$$\vartheta(t) = \frac{1}{\varkappa}\frac{\partial f(0,t)}{\partial t} + f(0,t) + \frac{\sigma_r^2}{2\varkappa^2}\left(1 - e^{-2\varkappa t}\right).$$

The process for $x(t) = r(t) - f(0,t)$ is

$$\left.\begin{aligned} dx(t) &= (y(t) - \varkappa x(t))\,dt + \sigma_r\,dW(t) \\ x(0) &= 0 \end{aligned}\right\}$$

where

$$y(t) = \frac{\sigma_r^2}{2\varkappa}\left(1 - e^{-2\varkappa t}\right).$$

We have already developed some methods for the Gaussian model, of which the Hull–White is a special case. Now, we present analytical formulae for some of the most important functions of the model.

5.5.1 Zero Coupon Bond in the Hull–White Model

The bond reconstitution formula in the Hull–White model follows from (5.21):

$$P(t,T,x(t)) = \frac{P(0,T)}{P(0,t)} \exp\left\{-x(t)\,G(t,T) - \frac{1}{2}y(t)\,G(t,T)^2\right\}$$

where

$$G(t,T) = \frac{1}{\varkappa}\left(1 - e^{-\varkappa(T-t)}\right).$$

5.5.2 Zero Coupon Bond Option in the Hull–White Model

The zero coupon bond option price is given by (5.25) or (5.26), the only term that needs to be computed is ν. First, we calculate $\sigma_P(t,u)$:

$$\sigma_P(t,u) = \sigma_r G(t,u) = \frac{\sigma_r}{\varkappa}\left(1 - e^{-\varkappa(u-t)}\right)$$

Next, the absolute difference:

$$|\sigma_P(u,T) - \sigma_P(u,T^*)| = \frac{\sigma_r}{\varkappa}\left(e^{-\varkappa(T-u)} - e^{-\varkappa(T^*-u)}\right)$$

Its square:

$$|\sigma_P(u,T) - \sigma_P(u,T^*)|^2 =$$

$$= \frac{\sigma_r^2}{\varkappa^2} \left(\exp\left\{-2\varkappa\left(T-u\right)\right\} - 2\exp\left\{-\varkappa\left(T-u+T^*-u\right)\right\} + \exp\left\{-2\varkappa\left(T^*-u\right)\right\}\right).$$

And finally integration gives

$$\nu = \int_t^T \left|\sigma_P\left(u,T^*\right) - \sigma_P\left(u,T\right)\right|^2 du =$$

$$= \frac{\sigma_r^2}{\varkappa^2} \int_t^T \left(\exp\left\{-2\varkappa\left(T-u\right)\right\} - 2\exp\left\{-\varkappa\left(T-u+T^*-u\right)\right\} + \exp\left\{-2\varkappa\left(T^*-u\right)\right\}\right) du =$$

$$=: \frac{\sigma_r^2}{\varkappa^2}\left(I_1+I_2+I_3\right).$$

We integrate the three terms separately:

$$I_1 := \int_t^T \exp\left\{-2\varkappa\left(T-u\right)\right\} du = \frac{1}{2\varkappa}\left[e^{-2\varkappa(T-u)}\right]_t^T = \frac{1}{2\varkappa}\left(1-e^{-2\varkappa(T-t)}\right),$$

$$I_2 := -2\int_t^T \exp\left\{-\varkappa\left(T-u+T^*-u\right)\right\} du = -\frac{2}{2\varkappa}\left[e^{-\varkappa(T+T^*-2u)}\right]_t^T =$$

$$= -\frac{2}{2\varkappa}\left(e^{-\varkappa(T^*-T)} - e^{-\varkappa(T^*+T-2t)}\right),$$

$$I_3 := \int_t^T \exp\left\{-2\varkappa\left(T^*-u\right)\right\} du = \frac{1}{2\varkappa}\left[e^{-2\varkappa(T^*-u)}\right]_t^T = \frac{1}{2\varkappa}\left(e^{-2\varkappa(T^*-T)} - e^{-2\varkappa(T^*-t)}\right).$$

5.5.3 Jamshidian Decomposition in the Hull–White Model

Recall from Section 4.1 that if there exists a process $x\left(t\right)$ so that $P\left(t,T\right) = P\left(t,T,x\left(t\right)\right)$ is a monotone and deterministic function of x, then we can apply Jamshidian's decomposition to compute the swaption price.

In particular, in the Hull–White model, $x\left(t\right) := r\left(t\right) - f\left(0,t\right)$ is such a process, and the zero coupon bond reconstitution formula is given by (5.21). So, the decomposition holds, and by (4.4) the swaption's price at time T_0 is:

$$V_{swaption}\left(T_0\right) = \sum_{i=0}^{n-1} \left(\delta N_i + cN_i\tau_i\right)\left(K_{i+1} - P\left(T_0,T_{i+1},x\left(T_0\right)\right)\right)^+.$$

This is the weighted sum of the pay-off functions of the zero coupon bond put options.

So, by the Black formula, the swaption price at time t is:

$$V_{swaption}(t) = \sum_{i=0}^{n-1} (\delta N_i + cN_i\tau_i) \left(P(t,T_0) K_{i+1} \Phi(-d_{-,i+1}) - P(t,T_{i+1}) \Phi(-d_{+,i+1})\right) \tag{5.27}$$

where K_i-s are produced as in (4.3),

$$d_{\pm,i+1} = \frac{\ln\left(P(t,T_{i+1}) / (KP(t,T_0))\right) \pm \nu_{i+1}/2}{\sqrt{\nu_{i+1}}}$$

$$\nu_{i+1} = \int_t^{T_0} |\sigma_P(u,T_{i+1}) - \sigma_P(u,T_0)|^2 \, du \,.$$

5.5.4 Hull–White Trinomial Tree

This section presents the trinomial tree approach proposed by Hull and White and is based on Hull & White (1993), Hull (1996), and Hull & White (1994*b*).

Since the Hull–White model is an equilibrium model and the mean $\vartheta(t)$ is chosen to perfectly match the yield curve, we want our tree to display this mean at every time step. Unfortunately, this cannot always be satisfied on a binomial tree, hence the need naturally arises to introduce a third child node which, as opposed to moving up or downward, represents moving exactly the amount of the drift. The building block of this model in the original setting of Hull & White (1993) looks like Figure 5.2.

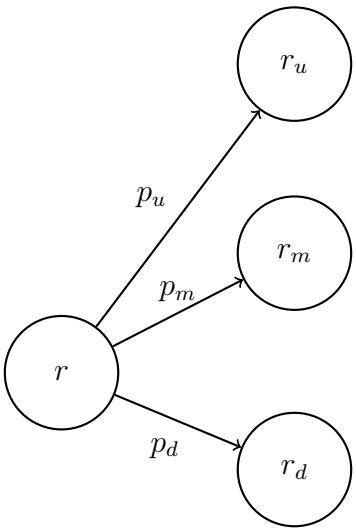


Figure 5.2: Building block of trinomial tree.

Following Hull & White (1993)'s approach, we construct such a tree in two steps: first, we build a centred, preliminary additive tree where the mean is exactly zero at all times, and then in the second step, we add a vertical adjustment, equalling the mean, at each time step.

Similarly to Nelson & Ramaswamy (1990)'s binomial tree, the grid in the preliminary tree is assumed to be equidistant both horizontally and vertically. Following the method and notation of Hull & White (1994*b*), let i and j be the index variables for the horizontal (time) and vertical (state) axes, respectively. Let Δt denote the equidistant step along the time axis and Δr along the vertical one. Let i be the index of the time variable such that $t_i = i\Delta t$ and j the state index such that $r_j = j\Delta r$. Note that in the middle of the tree, along $j = 0$, the rate is constant 0.[13] Figure 5.3 depicts an example.

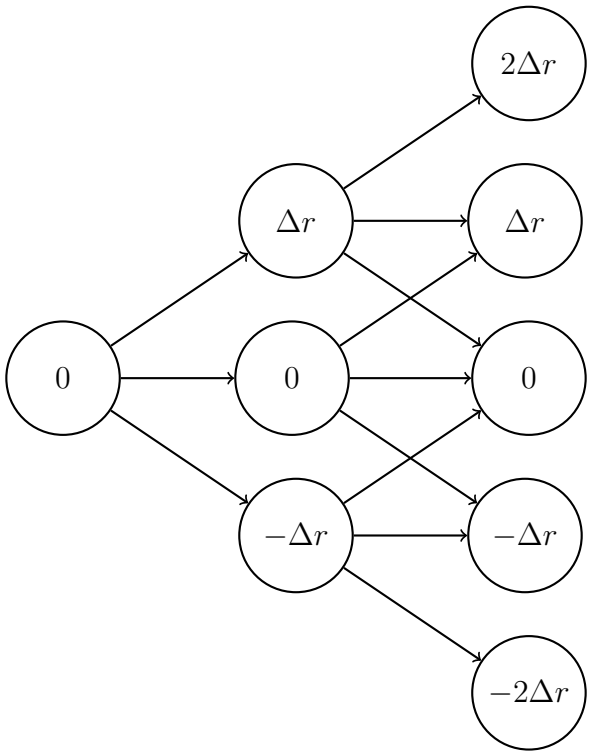


Figure 5.3: Preliminary trinomial tree.

If we take Δt as an exogenous variable – or hyperparameter –, it remains to determine Δr and the upward, middle and downward probabilities from each node. We have four variables, a probability constraint (the three must sum up to one), and we are left with three degrees of freedom. Using moment matching and suspecting that what we work

[13] Note that along every horizontal line, the rate is constant.

with are Gaussian variables, (thus the third moment must always be zero), we match the mean and the variance, and impose a restriction on Δr. Let $M \coloneqq -\varkappa\Delta t$ be the approximate mean, and $V \coloneqq \sigma^2\Delta t$ the approximate variance[14]. Following Hull & White (1993), we set $\Delta r = \sqrt{3V}$ as this is proven to minimise error.

Hull & White (1993) showed that the mean and variance are matched if we choose the probabilities from node (j, i) as

$$\begin{aligned} p_u &= \frac{1}{6} + \frac{j^2M^2 + jM}{2} \\ p_m &= \frac{2}{3} - j^2M^2 \\ p_d &= \frac{1}{6} + \frac{j^2M^2 - jM}{2}. \end{aligned}$$

It is apparent that these probabilities can fall outside of $[0, 1]$. To prevent this, Hull & White (1993) introduce alternative branching methods, as seen in Figure 5.4. In other words, we cut the tree at an upper index j_{max} and a lower index j_{min}. At the upper edge of the tree, at j_{max}, where the branching is in Figure 5.4a, the modified probabilities are as follows:

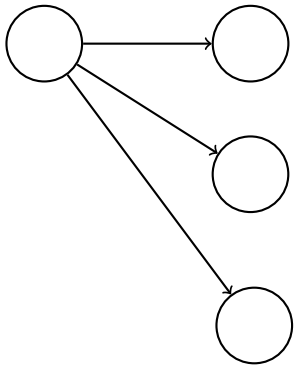
$$\begin{aligned} p_u &= \frac{7}{6} + \frac{j^2M^2 + 3jM}{2} \\ p_m &= -\frac{1}{3} - j^2M^2 - 2jM \\ p_d &= \frac{1}{6} + \frac{j^2M^2 + jM}{2}, \end{aligned}$$

and on the lower edge, at j_{min}, as in Figure 5.4b,

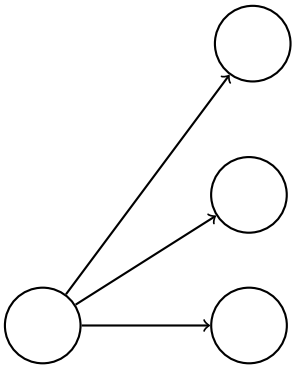
$$\begin{aligned} p_u &= \frac{1}{6} + \frac{j^2M^2 - jM}{2} \\ p_m &= -\frac{1}{3} - j^2M^2 + 2jM \\ p_d &= \frac{7}{6} + \frac{j^2M^2 - 3jM}{2}. \end{aligned}$$

Hull & White (1994*b*) shows that if we choose j_{max} as the smallest integer not less than $\frac{0.184}{M}$, and $j_{min} = -j_{max}$, then all probabilities are in $[0, 1]$. An alternative method would

[14]The exact mean is $e^{-\varkappa\Delta t} - 1$ and the exact variance is $\frac{\sigma^2\left(1-e^{-2\varkappa\Delta t}\right)}{2\varkappa}$, but using these values does not improve significantly the precision of calculations, according to Hull & White (1994*b*).



(a) Branching at the upper edge.



(b) Branching at the lower edge.

Figure 5.4: Alternative branching methods (Hull & White (1994*b*)).

be not to bother with cutting the tree, as asymptotically the results are the same, but with trees we rarely use fine enough grids to rely on convergence, moreover, it is desirable for the probabilities of the tree to be easy to comprehend. A tree cut at $j_{max} = 2$ is depicted in Figure 5.5.

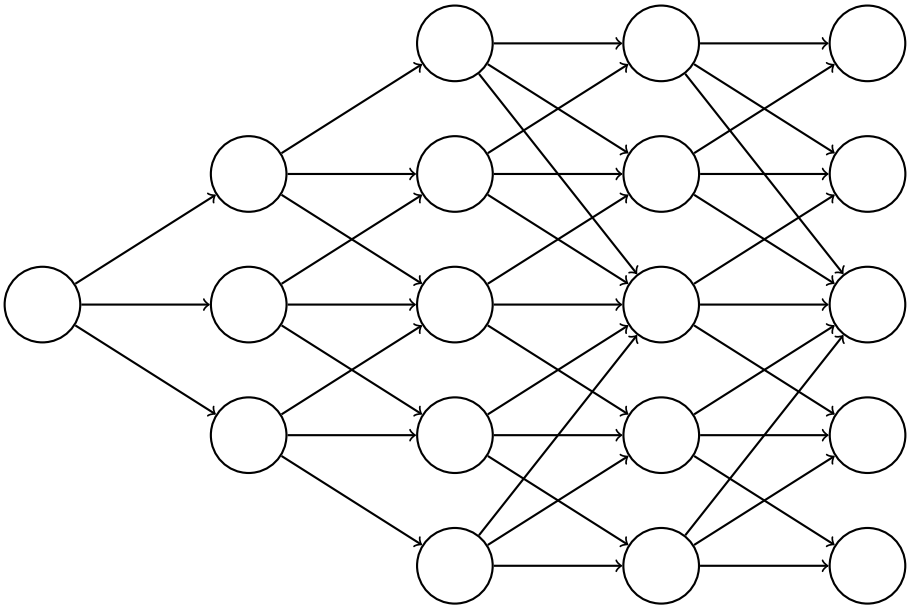


Figure 5.5: Trinomial tree cut at $j_{max} = 2$.

The first step of tree building is thus concluded, and it remains to adjust the tree at each time point on the grid to match the yield curve. One way to do this is to shift the vertical slice with the exact amount of $\vartheta(t)$, but as Hull (1996) indicates, due to discretisation, the mean of the tree will not exactly match that of the yield curve observed in the market. To resolve this issue, they present a lengthy method, using a fictional quantity $Q_{j,i}$ which

represents the pay-off 1 when node (j, i) is reached and 0 otherwise. The variable we are interested in is α_i, the shift over time i. Hull & White (1994*b*) derive the following recursion: α_0 is the Δt yield from the yield curve and $Q_{0,0} = 1$, and the rest can be computed from

$$Q_{j,m+1} = \sum_k Q_{k,m} p(k,j) \exp\{-(\alpha_m + k\Delta r)\Delta t\}$$

where $p(k,j)$ is the probability of moving from (k,m) to $(j,m+1)$, for all ks for which this is positive. To calculate α_m, the following recursion is used:

$$\alpha_m = \frac{-\ln \sum_{j=-n_m}^{n_m} Q_{j,m} e^{-j\Delta r \Delta t} - \ln P(0, m+1)}{\Delta t}.$$

For the whole derivation, see Hull & White (1994*b*). After we have calculated the shift for each time point on the grid, we are done constructing the Hull–White trinomial tree and follow the swaption pricing steps described in Section 5.3.3 to obtain the present value of the swaption.

Hull & White (1993)'s method became quite famous and can be used to price in a variety of no-arbitrage models. Hull himself provides examples for the Black & Karasinski (1991) or the Ho & Lee (1986) models in Hull (1996), and Leippold & Wiener (2003) extends the method to quadratic or affine models like the extended CIR.

5.5.5 Monte Carlo Methods in the Hull–White Model

We want to price a swaption with Monte Carlo under the Gaussian model. To do this, we follow the algorithm of generating Gaussian random variables, using that we compute trajectories of $x(t)$ under the risk-neutral measure, calculate the swaption's time T_0 value for each path, take a simple average and then discount back to time $t = 0$ with the zero coupon bond's price.

Note that we could generate paths of $r(t)$ as well, but $x(t)$ is simpler because it starts from 0.

To generate trajectories, we must choose a suitable discretisation scheme. As discussed in Section 4.3, there are several options to pick:

Explicit Euler–Maruyama Scheme

$$x_{i+1} = x_i + (y(t_i) - \varkappa(t_i)\, x_i)\, \Delta_i + \sigma_r(t_i) \sqrt{\Delta_i} Z_i \tag{5.28}$$

Linear Drift Euler–Maruyama Scheme

Similarly to the Vasicek model, consider the transformation $Y(t) \coloneqq e^{\varkappa t} x(t) - \int_0^t y(u)\, e^{\varkappa u}\, du$.

By Ito's lemma:

$$dY(t) = e^{\varkappa t} \sigma\, dW(t).$$

Apply an explicit Euler–Maruyama scheme scheme to Y:

$$Y_{i+1} = Y_i + e^{\varkappa t_i} \sigma \sqrt{\Delta_i} Z_i$$

where Z_i is a Gaussian variable with mean 0 and variance 1.

Now express $x(t)$ and $x(t + \Delta_i)$ as a function of $Y(t)$ and $Y(t + \Delta_i)$. Simple and straightforward calculus yields:

$$x_{i+1} = e^{-\varkappa \Delta_i} x_i + \left(1 - e^{-\varkappa \Delta_i}\right) \frac{y_i}{\varkappa} + e^{-\varkappa \Delta_i} \sigma \sqrt{\Delta_i} Z_i. \tag{5.29}$$

Implicit Euler–Maruyama Scheme

$$x_{i+1} = \frac{1}{1 + \varkappa(t_{i+1})\, \Delta_i} \left(x_i + y(t_{i+1})\, \Delta_i + \sigma_r(t_i) \sqrt{\Delta_i} Z_i \right). \tag{5.30}$$

Milstein Scheme

First, we have to calculate the partial derivatives of μ and σ, respectively, where $\mu(t, x) \coloneqq y(t) - \varkappa x(t)$ and $\sigma(t, x(t)) \coloneqq \sigma_r$.

$$\frac{\partial}{\partial t}\mu = \frac{d}{dt} y(t) = \frac{2\varkappa\sigma_r^2}{2\varkappa} e^{-2\varkappa t} = \sigma_r^2 e^{-2\varkappa t}$$
$$\frac{\partial}{\partial x}\mu = -\varkappa$$

and the rest is 0.

And thus the operators applied to μ and σ give:

$$\begin{aligned}\mathcal{L}_0\mu &= \sigma_r^2 e^{-2\varkappa t} - \varkappa\left(y\left(t\right) - \varkappa x\left(t\right)\right)\\ \mathcal{L}_1\mu &= -\varkappa\sigma_r\\ \mathcal{L}_0\sigma &= 0\\ \mathcal{L}_1\sigma &= 0.\end{aligned}$$

From here the Milstein scheme is:

$$\begin{aligned}x_{i+1} = x_i + \left(y\left(t_i\right) - \varkappa x_i\right)\Delta_i + \sigma_r\sqrt{\Delta_i}Z_{i,1} +\\ + \frac{1}{2}\left(\sigma_r^2 e^{-2\varkappa t_i} - \varkappa\left(y\left(t_i\right) - \varkappa x_i\right)\right)\Delta_i^2 - \varkappa\sigma_r\sqrt{\frac{\Delta_i^3}{3}}Z_{i,2}\end{aligned} \tag{5.31}$$

where the correlation of standard Gaussian variables $Z_{i,1}$ and $Z_{i,2}$ is $\sqrt{\frac{3}{4}}$.

5.6 The Cox–Ingersoll–Ross Model

One major drawback of the Hull–White model is that it allows negative interest rates. Although recently the extension of interest rate models to negative rates has become an exciting and relevant field of research (Antonov et al. (2015), Brigo & Mercurio (2001), Di Francesco & Kamm (2022)), for many years it was a highly undesirable characteristic, ought to be avoided if possible. Hence, a new branch of short rate models emerged to exclude the possibility of negative values. One group of such models is called *affine short rate models*[15], where the most general form for the short rate dynamics is

$$dr\left(t\right) = \varkappa\left(t\right)\left(\vartheta\left(t\right) - r\left(t\right)\right)dt + \sigma_r\left(t\right)\sqrt{\alpha\left(t\right) + \beta\left(t\right)r\left(t\right)}\,dW\left(t\right). \tag{5.32}$$

Though formally we allow α and β to be time-dependent, and there exist some generalisations (e.g. Duffie & Kan (1996)) which can handle these cases, the literature is almost always only concerned with the constant case. In particular, when $\varkappa$, ϑ and σ_r are also flat, the SDE is as follows:

$$dr\left(t\right) = \varkappa\left(\vartheta - r\left(t\right)\right)dt + \sigma_r\sqrt{\alpha + \beta r\left(t\right)}\,dW\left(t\right). \tag{5.33}$$

[15]It is called affine because the square of the diffusion term is, whenever σ_r is constant, an affine function of the short rate.

To ensure that the square root term never reaches zero, in other words, the SDE is well defined and remains stochastic, the parameters must satisfy the *Feller condition*[16]:

$$\varkappa\beta\left(\vartheta+\frac{\alpha}{\beta}\right)>\frac{1}{2}\beta^2\sigma_r^2$$

Remark. Note that the time-depending $\vartheta(t)$, $\alpha=1$, $\beta=0$ returns the Hull–White model. If ϑ is also constant, we get the Vasicek model. In these cases, the Feller condition is useless and always satisfied, indicating that the HW and Vasicek SDEs are always well defined.

The simplest case where the diffusion term is truly a function of $r(t)$ was first proposed by Cox et al. (1985) and is called the Cox–Ingersoll–Ross model or, in short, the CIR model. Its short rate SDE is, with $\alpha=0,\beta=1$, all other terms constant:

$$dr(t)=\varkappa(\vartheta-r(t))\,dt+\sigma\sqrt{r(t)}\,dW(t). \tag{5.34}$$

As with all other interest rate models, we are first and foremost interested in the zero coupon bond formula. To derive it for the CIR model, we take a short detour and have a look at the so-called Affine Term Structure models where the zero coupon bond is a transformation – more specifically, the exponential – of an affine function of the short rate[17], in other words, it can be written as

$$P(t,T)=\exp\{A(t,T)-B(t,T)\,r(t)\}. \tag{5.35}$$

5.6.1 Term Structure Equation and Affine Term Structure Models

This section is based on Medvegyev (2014)[18]. Suppose first that we have a model describing the dynamics of the short rates:

$$dr(t)=\mu(t,r)\,dt+\sigma(t,r)\,dW(t). \tag{5.36}$$

Our aim is to derive a formula for the zero coupon bond price.

[16] Feller (1951).

[17] Note that, by this definition, affine short rate and affine term structure models not necessarily coincide. Duffie & Kan (1996) showed, though, that affine short models are always special cases of affine term structure models. The opposite implication does not hold.

[18] Pp. 374-382.

Recall that in the single factor Black & Scholes (1973) world we have three tradable securities: the numéraire, an underlying, whose dynamics is known, and a derivative whose price depends on the underlying. From this set-up, under a few other assumptions, the most important being the no-arbitrage condition, we can derive the Black–Scholes–Merton equation, a second-order partial differential equation for the price of the derivative.

Now consider the zero coupon bond. It is a tradable security with price depending on the short rate. Suppose we have a numéraire as well, with the same dynamics as in the Black–Scholes world, namely, the value increment being the previous value multiplied by the *short rate*. Under the no-arbitrage condition, though the short rate itself not being tradable, we might suspect that the zero coupon bond price $P(t,T)$ can satisfy something very similar to the BSM-equation, the *Term Structure Equation*:

$$\frac{\partial}{\partial t}P(t,T,r)+\mu(t,r)\frac{\partial}{\partial r}P(t,T,r)+\frac{1}{2}\sigma^2(t,r)\frac{\partial^2}{\partial r^2}P(t,T,r)=rP(t,T,r)$$
$$P(T,T,r)=1. \tag{5.37}$$

Note that on the right-hand side, we have the same r. It was a constant in the Black–Scholes model, and now it is a variable.

Now, analogously to the Black–Scholes derivation, let $X(t) \coloneqq P(t,T,r(t))$, and let $D(t) \coloneqq \exp\left\{-\int_0^t r(s)\,ds\right\}$ be the discount factor, the reciprocal of the numéraire. It is evident that

$$dD(t)=-r(t)D(t)\,dt\,. \tag{5.38}$$

By Ito's lemma:

$$dX(t)=\left(\frac{\partial}{\partial t}P(t,T,r)+\mu(t,r)\frac{\partial}{\partial r}P(t,T,r)+\frac{1}{2}\sigma^2(t,r)\frac{\partial^2}{\partial r^2}P(t,T,r)\right)dt+$$
$$+\sigma(t,r)\frac{\partial}{\partial r}P(t,T,r)\,dW(t)\,. \tag{5.39}$$

From the integration-by-parts formula, using that D is deterministic, and consequently

quadratic cross-variation is zero, and the Term Structure Equation:

$$\begin{aligned}
[X(s)D(s)]_t^u &= \int_t^u X(s)\,dD(s) + \int_t^u D(s)\,dX(s) = \\
&= \int_t^u -r(s)P(s,T,r)D(s)\,ds + \\
&+ \int_t^u D(s)\left(\frac{\partial}{\partial s}P(s,T,r) + \mu(s,r)\frac{\partial}{\partial r}P(s,T,r) + \frac{1}{2}\sigma^2(s,r)\frac{\partial^2}{\partial r^2}P(s,T,r)\right)ds + \\
&+ \int_t^u \sigma(s,r)\frac{\partial}{\partial r}P(s,T,r)D(s)\,dW(s) = \\
&= \int_t^u \sigma(s,r)\frac{\partial}{\partial r}P(s,T,r)D(s)\,dW(s).
\end{aligned}$$

We see that under some mild assumptions[19] the discounted process for X is a martingale, so $X(t) = \exp\left\{\int_0^t r(s)\,ds\right\}\mathbb{E}_t\left[\exp\left\{-\int_0^T r(s)\,ds\right\}1\right]$. So X is indeed the zero coupon bond price. Hence, to give a formula for $P(t,T)$, we have to solve the Term Structure Equation (5.37). The PDE is not always solvable, so we often have to make some assumptions about the class of functions of the solution. One such restriction is that the zero coupon bond is the exponential of an affine function of the short rate, and thus it can be written as

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

In this case, after dividing by $P(t,T)$, the TSE is the following:

$$\left.\begin{aligned}
&\tfrac{\partial}{\partial t}A(t,T) - \tfrac{\partial}{\partial t}B(t,T)r - \mu(t,r)B(t,T) + \tfrac{1}{2}\sigma(t,r)^2 B(t,T)^2 = r \\
&A(T,T) = 0, \quad B(T,T) = 0
\end{aligned}\right\} \tag{5.40}$$

We saw that in affine short rate models $\mu(t,r) := \gamma(t)r + \delta(t)$ and $\sigma(t,r)^2 := \alpha(t) +$

[19] $P(t,T,r)$ must be uniformly bounded (Medvegyev (2014)).

$\beta(t) r$[20]. Now, by the TSE, using that r is a variable, A and B must satisfy the following:

$$\left.\begin{array}{ll} \frac{\partial}{\partial t} A(t,T) - \delta(t) B(t,T) + \frac{1}{2}\alpha(t) B(t,T)^2 & = 0 \\ \frac{\partial}{\partial t} B(t,T) + \gamma(t) B(t,T) - \frac{1}{2}\beta(t) B(t,T)^2 & = -1 \\ A(T,T) = 0, \quad B(T,T) = 0 & \end{array}\right\} \tag{5.41}$$

This is a Riccati ordinary differential equation system and is often easy to solve. The CIR model is the special case $\alpha(t) = 0$, $\beta(t) = \sigma^2$, $\gamma(t) = -\varkappa$ and $\delta(t) = \varkappa\vartheta$. [21]

5.6.2 Zero Coupon Bond in the CIR Model

To obtain the zero coupon bond reconstitution formula, we will solve the (5.41) system which translates into

$$\frac{\partial}{\partial t} A(t,T) - \varkappa\vartheta B(t,T) = 0 \tag{5.42}$$

$$\frac{\partial}{\partial t} B(t,T) - \varkappa B(t,T) - \frac{1}{2}\sigma^2 B(t,T)^2 = -1 \tag{5.43}$$

$$A(T,T) = 0, \quad B(T,T) = 0$$

The second equation is a Riccati ODE, and after solving that, we can simply integrate to solve the first one. A Riccati equation requires a particular solution, which we will search in constant form[22]. Let y^* be a constant particular solution to (5.43). Then from the quadratic equation,

$$0 = \varkappa y^* + \frac{1}{2}\sigma^2 y^{*2}$$

$$y^* = \frac{-\varkappa - \sqrt{\varkappa^2 + 2\sigma^2}}{\sigma^2} =: \frac{-\varkappa - \gamma}{\sigma^2}.$$

Let $y := y^* + u$ be the general solution to (5.43). By simple calculus,

$$y' = y^{*\prime} + u' = u' = \varkappa(y^* + u) + \frac{1}{2}\sigma^2 (y^* + u)^2 - 1 =$$

[20] Note that the α and β here not necessarily equals those in the previous chapter, we have to multiply those by σ^2.

[21] Note that the Vasicek and the Hull–White model can also be written in this form, meaning that they are also affine term structure models. Their ODE systems are linear as well and, of course, return the same result for the zero coupon bond that we have obtained in earlier chapters.

[22] Following Medvegyev (2014)

$$= \underbrace{\varkappa y^* + \frac{1}{2}\sigma^2 y^{*2} - 1}_{0} + \varkappa u + \sigma^2 y^* u + \frac{1}{2}\sigma^2 u^2 =$$
$$= \varkappa u - (\varkappa + \gamma) u + \frac{1}{2}\sigma^2 u^2 = -\gamma u + \frac{1}{2}\sigma^2 u^2.$$

We introduce a new variable $z \coloneqq \frac{1}{u}$, $z' = -\frac{u'}{u^2}$. The ODE for z:

$$z' = \gamma z - \frac{1}{2}\sigma^2.$$

This is a first-order linear ODE. The general solution by the Cauchy formula is

$$z(t) = e^{\gamma t}\left(z_0 - \int_0^t e^{-\gamma s}\frac{1}{2}\sigma^2\, ds\right) = e^{\gamma t} z_0 + \frac{\sigma^2}{2\gamma}\left(1 - e^{\gamma t}\right).$$

From the boundary condition $B(T,T) = y(T) = 0$, we have

$$z(T) = -\frac{1}{y^*(T)} = \frac{\sigma^2}{\gamma + \varkappa} = e^{\gamma T} z_0 + \frac{\sigma^2}{2\gamma}\left(1 - e^{\gamma T}\right)$$
$$z_0 = e^{-\gamma T}\left(\frac{\sigma^2}{\gamma + \varkappa} - \frac{\sigma^2}{2\gamma}\right) + \frac{\sigma^2}{2\gamma}.$$

So the solution for z is

$$z(t) = e^{-\gamma(T-t)}\left(\frac{\sigma^2}{\gamma + \varkappa} - \frac{\sigma^2}{2\gamma}\right) + e^{\gamma t}\frac{\sigma^2}{2\gamma} + \frac{\sigma^2}{2\gamma} - \frac{\sigma^2}{2\gamma}e^{\gamma t} =$$
$$= e^{-\gamma(T-t)}\left(\frac{\sigma^2}{\gamma + \varkappa} - \frac{\sigma^2}{2\gamma}\right) + \frac{\sigma^2}{2\gamma} = \frac{e^{\gamma t} z_0 2\gamma + \sigma^2\left(1 - e^{\gamma t}\right)}{2\gamma}.$$

Taking reciprocal and adding the particular solution, we obtain the general solution for $B(t,T)$:

$$B(t,T) = y^* + \frac{2\gamma e^{-\gamma t}}{z_0 2\gamma + \sigma^2\left(e^{-\gamma t} - 1\right)} = \frac{y^* z_0 2\gamma + y^*\sigma^2\left(e^{-\gamma t} - 1\right) + 2\gamma e^{-\gamma t}}{e^{-\gamma T}\left(\frac{\sigma^2}{\varkappa+\gamma}2\gamma - \sigma^2\right) + \sigma^2\left(e^{-\gamma t} - 1\right)} =$$
$$= \frac{y^* e^{-\gamma T}\left(\frac{\sigma^2}{\varkappa+\gamma}2\gamma - \sigma^2\right) + y^*\sigma^2 + y^*\sigma^2 e^{-\gamma t} - y^*\sigma^2 + 2\gamma e^{-\gamma t}}{e^{-\gamma T}\left(\frac{\sigma^2}{\varkappa+\gamma}2\gamma - \sigma^2\right) + \sigma^2\left(e^{-\gamma t} - 1\right)} =$$
$$= \frac{-e^{-\gamma T}2\gamma + (\varkappa + \gamma)\, e^{-\gamma T} - e^{-\gamma t}(\varkappa + \gamma) + 2\gamma e^{-\gamma t}}{e^{-\gamma T}\left(\frac{\sigma^2}{\varkappa+\gamma}2\gamma - \sigma^2\right) + \sigma^2\left(e^{-\gamma t} - 1\right)} =$$

$$= \frac{(\gamma-\varkappa)(\varkappa+\gamma)\left(1-e^{-\gamma(T-t)}\right)}{\sigma^2\left((\varkappa+\gamma)-e^{-\gamma(T-t)}(\varkappa-\gamma)\right)} = \frac{\gamma^2-\varkappa^2}{\sigma^2}\frac{\left(e^{\gamma(T-t)}-1\right)}{(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)+2\gamma} =$$
$$= \frac{2\left(e^{\gamma(T-t)}-1\right)}{(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)+2\gamma}. \tag{5.44}$$

We show that

$$A(t,T) = \frac{\varkappa\vartheta}{\sigma^2}\left((\varkappa+\gamma)(T-t) - 2\ln\left(\frac{2\gamma+(\varkappa+\vartheta)\left(e^{\gamma(T-t)}-1\right)}{2\gamma}\right)\right). \tag{5.45}$$

Differentiating gives

$$\frac{\partial}{\partial t}A(t,T) = \frac{\varkappa\vartheta}{\sigma^2}\left(-(\varkappa+\gamma)+2\frac{2\gamma\frac{\varkappa+\gamma}{2\gamma}\gamma e^{\gamma(T-t)}}{2\gamma+(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)}\right) =$$
$$= \frac{2\varkappa\vartheta}{\sigma^2}\left(\frac{\left(\gamma-\frac{\varkappa+\gamma}{2}\right)(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)-\gamma(\varkappa+\gamma)+\gamma(\varkappa+\gamma)}{2\gamma+(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)}\right) =$$
$$= \frac{\varkappa\vartheta}{\sigma^2}\frac{(\gamma-\varkappa)(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)}{2\gamma+(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)} = \varkappa\vartheta\frac{\gamma^2-\varkappa^2}{\sigma^2}\frac{e^{\gamma(T-t)}-1}{2\gamma+(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)} =$$
$$= \varkappa\vartheta\frac{2\left(e^{\gamma(T-t)}-1\right)}{2\gamma+(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)} = \varkappa\vartheta B(t,T)$$

which satisfies (5.42).

A more wide-spread formula for the zero coupon bond in the CIR model is $P(t,T) = A_0(t,T)\,e^{-B(t,T)r(t)}$ where

$$A_0(t,T) = e^{A(t,T)} = \left(\frac{e^{\frac{(\varkappa+\gamma)(T-t)}{2}}2\gamma}{2\gamma+(\varkappa+\gamma)\left(e^{\gamma(T-t)}-1\right)}\right)^{\frac{2\varkappa\vartheta}{\sigma^2}}. \tag{5.46}$$

5.6.3 Zero Coupon Bond Option in the CIR Model

An undeniable advantage of the Cox–Ingersoll–Ross model is that there exist analytical formulae to price options on bonds. The recipe is the same as in the Black world: find the distribution of the zero coupon bond, then calculate the $r(T) = r^*$ value of the short rate for which the option is at the money at expiry. Finally, with the help of the cumulative distribution function (CDF) of the bond's distribution – evaluated at two cleverly chosen points – and this r^*, apply the usual formula: $P(t,T^*)\cdot CDF(d_+) - P(t,T)\,K\cdot CDF(d_-)$, where t is today's time, T is the expiration of the option, T^* is the maturity of the underlying bond, K is the strike, and $d_\pm$ are functions of the model parameters, including

r^*.

Deriving the distribution of the short rate or the zero coupon bond in the CIR model is beyond the scope of this dissertation, nevertheless Cox et al. (1985) showed that they both follow non-central chi-squared distributions. They derived the bond option price as well for call options on zero coupon bonds. It is as follows:

$$\begin{aligned}\mathbf{ZBC}\left(r,t,T,T^*,K\right) &= P\left(r,t,T^*\right)\chi^2\left(2r^*\left(\phi+\psi+B\left(T,T^*\right)\right);\frac{4\varkappa\vartheta}{\sigma^2};\frac{2\phi^2 re^{\gamma(T-t)}}{\phi+\psi+B\left(T,T^*\right)}\right)+\\ &\quad -KP\left(r,t,T\right)\chi^2\left(2r^*\left(\phi+\psi\right);\frac{4\varkappa\vartheta}{\sigma^2};\frac{2\phi^2 re^{\gamma(T-t)}}{\phi+\psi}\right) \qquad (5.47)\\ \gamma &= \sqrt{\varkappa^2+2\sigma^2}\\ \phi &= \frac{2\gamma}{\sigma^2\left(e^{\gamma(T-t)}-1\right)}\\ \psi &= \frac{\varkappa+\gamma}{\sigma^2}\\ r^* &= \frac{\ln\left(\frac{A_0(t,T^*)}{K}\right)}{B\left(t,T^*\right)}\end{aligned}$$

where $\chi^2\left(x,df,nc\right)$ denotes the cumulative distribution function of the non-central chi-squared distribution at x with degree of freedom df and non-centrality parameter nc. Note that unlike in the Gaussian model, the option price is a function of the current short rate $r\left(t\right)$. This is in part because we have not applied a change of variables, in part because the CIR model is an equilibrium model, and its parameters cannot be chosen so that the implied yield curve perfectly fits the one observed in the market.

To apply the Jamshidian decomposition on swaptions, we need a formula for the put option as well. It is easily obtained from the put-call parity, an identity independent of the model choice:

$$\mathbf{ZBP}\left(r,t,T,T^*,K\right) = \mathbf{ZBC}\left(r,t,T,T^*,K\right) - P\left(r,t,T^*\right) + KP\left(r,t,T\right). \qquad (5.48)$$

Now we have all the tools to apply Jamshidian's trick to the swaption in the CIR model. In this case, the variable x will be the short rate r itself. We have to solve (4.2) for a short rate r^*[23], then use the decomposition and price the zero coupon bond put options to obtain the swaption price.

[23]Note that it is **not** the same as r^* in the bond option formula.

5.6.4 Binomial Tree for the CIR Model

This section presents a binomial tree approach applied to the CIR model and is based on Nelson & Ramaswamy (1990).

Unfortunately, the method presented in Section 5.3.3 for the Vasicek model does not work for models where the diffusion term of the underlying process is not constant, as is the case with the CIR model (Nelson & Ramaswamy (1990)). We need to introduce an intermediary step to construct a simple tree for a variable X that follows a constant diffusion process, and then transform it back with $f^{-1}(r) \coloneqq X$ with f being a regular enough function[24].

In particular, let $X = \frac{2\sqrt{r}}{\sigma}$. By Ito's lemma,

$$\begin{aligned} dX(t) &= \frac{1}{\sqrt{r}\sigma} dr - \frac{1}{4\sqrt{r^3}\sigma} d[r] = \\ &= \frac{1}{\sqrt{r}\sigma} \varkappa(\vartheta - r) dt + \frac{1}{\sqrt{r}\sigma} \sqrt{r}\sigma \, dW(t) - \frac{1}{4\sqrt{r^3}\sigma} r\sigma^2 dt = \\ &= \left(\frac{2\varkappa\vartheta}{X(t)\sigma^2} - \frac{1}{2X(t)} - \frac{\varkappa X(t)}{2} \right) dt + dW(t) \end{aligned}$$

which has the desired property. We then construct a tree for X in the same way as in Section 5.3.3, then use the transformation $r = \frac{\sigma^2 X^2}{4}$ and the probability of moving upward from a node r

$$p(r) \coloneqq \frac{\varkappa(\vartheta - r)\Delta + r - r_-}{r_+ - r_-}$$

where r_+ and r_- are the children nodes up and down from node r respectively. Using this probability, the local drift of the r process will be matched exactly (Nelson & Ramaswamy (1990)). The authors also showed that the Cox–Ross–Rubinstein model for the log-normal equity prices is a special application of this method with $f(X) \coloneqq e^{\sigma X}$. We do not trouble ourselves with proving that the transforming function in the CIR model satisfies the regularity conditions, for a comprehensive and rigorous approach on regularity issues, see Nelson & Ramaswamy (1990).

The rest of the pricing with a tree goes exactly as described in Section 5.3.3, with the help of discount factors, working our way back on the time grid.

[24] For precise conditions, see Nelson & Ramaswamy (1990).

5.6.5 Monte Carlo Methods in the CIR model

To run the Monte Carlo simulation in the CIR model, we must generate trajectories. All the usual techniques work, but higher-order schemes require a little more computation.

The Explicit Euler–Maruyama Scheme

The simplest discretisation is:

$$r_{i+1} = r_i + \varkappa\left(\vartheta - r_i\right)\Delta_i + \sigma\sqrt{r_i}\sqrt{\Delta_i}Z_i \tag{5.49}$$

where Z_i is a standard Gaussian random variable.

Linear Drift Euler–Maruyama Scheme

$$r_{i+1} = e^{-\varkappa\Delta_i}r_i + \left(1 - e^{-\varkappa\Delta_i}\right)\theta + e^{-\varkappa\Delta_i}\sigma\sqrt{r_i}\sqrt{\Delta_i}Z_i. \tag{5.50}$$

The Milstein Scheme

First, we establish that $\mu\left(t, r\left(t\right)\right) \coloneqq \varkappa\vartheta - \varkappa r\left(t\right)$ and $\sigma\left(t, r\left(t\right)\right) \coloneqq \sigma\sqrt{r\left(t\right)}$. Now we have to apply the $\mathcal{L}_0$ and $\mathcal{L}_1$ operators to each of them:

$$\mathcal{L}_0\mu = \left(\frac{\partial}{\partial t} + \mu\left(t,r\right)\frac{\partial}{\partial r} + \frac{1}{2}\sigma\left(t,r\right)^2\frac{\partial^2}{\partial r^2}\right)\mu\left(t,r\right) = 0 - \mu\left(t,r\right)\varkappa + 0 = -\varkappa^2\left(\vartheta - r\left(t\right)\right)$$

$$\mathcal{L}_1\mu = \sigma\left(t,r\right)\frac{\partial}{\partial r}\mu\left(t,r\right) = -\varkappa\sigma\sqrt{r\left(t\right)}$$

$$\mathcal{L}_0\sigma = \left(\frac{\partial}{\partial t} + \mu\left(t,r\right)\frac{\partial}{\partial r} + \frac{1}{2}\sigma\left(t,r\right)^2\frac{\partial^2}{\partial r^2}\right)\sigma\left(t,r\right) =$$

$$= 0 + \varkappa\left(\vartheta - r\left(t\right)\right)\frac{\sigma}{2\sqrt{r\left(t\right)}} - \frac{1}{2}\sigma^2 r\left(t\right)\frac{1}{4}\frac{\sigma}{\sqrt{r\left(t\right)^3}} = \frac{\varkappa\sigma}{2}\left(\frac{\vartheta}{\sqrt{r\left(t\right)}} - \sqrt{r\left(t\right)}\right) - \frac{1}{8}\frac{\sigma^3}{\sqrt{r\left(t\right)}}$$

$$\mathcal{L}_1\sigma = \sigma\left(t,r\right)\frac{\partial}{\partial r}\sigma\left(t,r\right) = \sigma\sqrt{r\left(t\right)}\sigma\frac{1}{2\sqrt{r\left(t\right)}} = \frac{\sigma^2}{2}$$

Inserting these into (4.15), we get

$$r_{i+1} = r_i + \varkappa\left(\vartheta - r_i\right)\Delta_i + \sigma\sqrt{r_i}\sqrt{\Delta_i}Z_{i,1} - \frac{\Delta_i^2}{2}\varkappa^2\left(\vartheta - r_i\right) +$$

$$- \varkappa\sigma\sqrt{r_i}\sqrt{\frac{\Delta_i^3}{3}}Z_{i,2} +$$

$$+\left(\frac{\varkappa\sigma}{2}\left(\frac{\vartheta}{\sqrt{r\left(t\right)}}-\sqrt{r\left(t\right)}\right)-\frac{1}{8}\frac{\sigma^{3}}{\sqrt{r\left(t\right)}}\right)\left(\Delta_{i}Z_{i,1}\sqrt{\Delta_{i}}-\sqrt{\frac{\Delta_{i}^{3}}{3}}Z_{i,2}\right)+$$
$$+\frac{1}{2}\sigma^{2}\left(\frac{1}{2}Z_{i,1}^{2}\Delta_{i}-\frac{1}{2}\Delta_{i}\right). \tag{5.51}$$

where $Z_{i,1}$ and $Z_{i,2}$ are standard Gaussian variables with correlation $\frac{3}{4}$.

Chapter 6

Analysis

After examination of the theory of swaption pricing, we implemented the three pricing techniques mentioned above. This section details the implementation, performance and sensitivity of each method.

6.1 Implementation

The project was conducted in a PyCharm environment under Python 3.10. and in Jupyterlab notebooks.

6.1.1 Instruments and Models

We implemented classes of different instruments: a yield curve, a swap, a swaption; and the three models: the Vasicek, the Hull–White and the CIR.

Yield Curve

The yield curve object has two arguments: an array of dates and a corresponding array of yields. It is assumed to be piecewise flat and left-continuous.

It serves two purposes. Its first date is assumed to be today's date, although since all data are fictional, this is only a convention, used in each of the models as the pricing date $t = 0$. Second, in the Hull–White model, it gives the exogenous "market-observed" yield curve, used in simulation, discounting, and calibrating the tree.

The yield curve object has functions to calculate today's discount factor $P(0, t)$ and forward discount factor $P(0, t, T)$. If t is not in the array of dates of the yield curve,

the function still returns the correct factor using the piecewise flat and left-continuous property.

Swap

The swap object has many arguments as can be seen in Table 6.1. It represents a payer swap with flat coupon term structure, zero spread and the same payment schedule for both legs. The first notional is only a nominal value and is assumed throughout the analysis to be 1. The notional increment is constant over the tenor, so it excludes roller-coaster swaps, but allows for accreters and amortisers. A negative value corresponds to the latter. The coupon is also constant and is given as fraction of the notional. Frequency refers to monthly, quarterly, semi-annual or annual payment dates, and the tenor is the number of these payments. The start date of the swap means the beginning of the first accrual period.

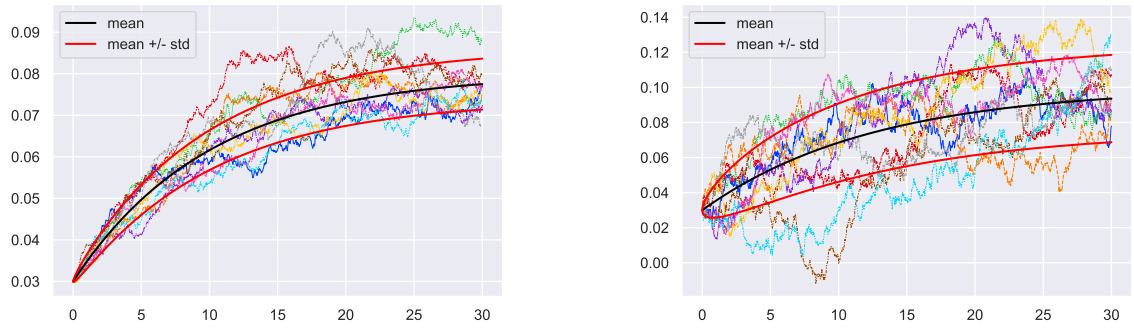
first notional:	the first notional N_0
notional increment:	the uniform increment in notionals δN
coupon:	coupon of the fixed leg
start date:	the start of the first accrual period
tenor:	number of accrual periods
freq:	frequency of payments
yc:	yield curve object
model:	model object, either Vasicek, CIR or Hull–White

Table 6.1: Swap characteristics.

The instrument can exist on its own, but takes two other arguments: a yield curve object, mainly to mark today's date, and a model framework to help future pricing. It has an annuity and a swap rate calculating method; both can handle the case where the pricing date is not today, but in this case, of course, we have to use the underlying model's assumptions about rate dynamics.

Swaption

The swaption object only takes the underlying swap as an argument and sets its expiry date as the start date of the swap, since we do not have the means to price mid-curve swaptions. It has three methods, each to price the swaption: the Jamshidian pricer, a Monte Carlo engine and a lattice tree, all three heavily depending on the underlying model.



(a) CIR (b) Vasicek

Figure 6.1: CIR and Vasicek trajectories.

Vasicek and CIR model

The Vasicek and CIR model classes take five arguments: $\varkappa, \vartheta, \sigma, r_0$ and a yield curve object to extract today's date. They have a few functions to calculate different model-specific metrics like $A(t,T), B(t,T)$, the bond price and the bond option prices, and a function to get the theoretical mean and standard deviation of the short rate at any time. They have some simulation methods as well: a tree constructor and some Monte Carlo path simulator engines. Figure 6.1 presents a few possible trajectories of short rates for the CIR and Vasicek models with the theoretical mean and a standard deviation error bar. Notice that the variance of the Vasicek is significantly higher, and there are a few instances when the rate becomes negative.

Hull–White model

The Hull–White model takes only three arguments: $\varkappa, \sigma$ and a yield curve object. It also has model-specific metrics like $G(t,T), y(t)$, the bond price and the bond option prices, and of course some simulation methods: a tree and Monte Carlo engines.

6.1.2 Jamshidian Pricer

The Jamshidian pricer calculates the value of the payer swaption by following the method in Section 4.1. A pseudo code of the algorithm is in Algorithm 1.

Algorithm 1 Jamshidian algorithm

```
input: swap
find x* so that (4.2) is satisfied
for i = 1 to n do
    K_i ← (4.3)
    if HW then
        ZBP_i ← (5.26)
    else if VASICEK then
        ZBP_i ← (5.13)
    else
        ZBP_i ← (5.48)
    end if
end for
swaption ← (4.5)
return swaption
```

6.1.3 Monte Carlo Pricer

The Monte Carlo engine also computes the present value of the swaption. It takes four arguments: the number of simulations M, an integer n representing the finesse of the time grid (i.e., the number of equidistant steps until the start date of the swap) set to 100 as default, a *method* argument that specifies which discretisation scheme to use – Euler, Linear Drift (default) or Milstein – and a Boolean *converge*, which, if true, returns the swaption price in each simulation, otherwise it only returns the mean. The pseudocode in Algorithm 2 demonstrates the algorithm.

6.1.4 Tree Pricer

The lattice tree also prices the swaption today. It does not take any arguments, but creates a time grid up to the last payment date of the swap, with time step equalling the length of the accrual period – this means that it can only price swaptions with a regular payment schedule, and the expiry date *must* fall on the grid. There are three slightly different approaches[1] to build the tree, based on the model we are using. Algorithm 3 presents these.

[1] See Sections 5.3.3, 5.5.4 and 5.6.4.

Algorithm 2 Monte Carlo algorithm

input: swap, M, n, method, converge
timeGrid $\leftarrow$ **linspace**(today, exDate, num $= n$)
if HW **then**
  $r[0] = 0$
else
  $r[0] = r_0$
end if
for $j = 1$ to M **do**
  price $\leftarrow 0$
  S $\leftarrow$ [ ]
  for $i = 1$ to **len**(timeGrid) **do**
    if method **is** Euler **then**
      $r[i] \leftarrow$ (5.49) **or** (5.15) **or** (5.28)
    else if method **is** Linear Drift **then**
      $r[i] \leftarrow$ (5.50) **or** (5.16) **or** (5.29)
    else if method **is** Milstein **then**
      $r[i] \leftarrow$ (5.51) **or** (5.17) **or** (5.31)
    end if
  end for
  swaption $\leftarrow$ (4.1) **with** $r(T_0) = r[-1]$ **and** ((5.21) **or** (5.11) **or** ((5.45) **and** (5.44))
  price $\leftarrow$ price + swaption
  if converge **then**
    S.append swaption
  end if
end for
if converge **then**
  return S
else
  return price/M
end if

Algorithm 3 Tree algorithm

input: swap
timeGrid $\leftarrow$ **linspace**(today, endDate, dt = frequency)
if HW **then**
    construct first tree as in Section 5.5.4 by forward induction
    construct rTree and probTree as in Section 5.5.4 by forward induction
else if VASICEK **then**
    construct rTree and probTree as in Section 5.3.3
else
    construct first tree as in Section 5.6.4 by forward induction
    construct rTree and probTree as in Section 5.6.4 by forward induction
end if
construct tree of discount factors P_1 from rTree
P $\leftarrow [P_1]$
for $i = 1$ to tenor **do**
    $P_{prev} = P[-1]$
    construct P_{curr} from P_{prev}, probTree and P_1 **with** (5.14)
    P.append P_{curr}
end for
for each node in grid at exDate **do**
    calculate swaption value
end for
from exDate, use backward induction and probTree and P_1 to today's node
return the value at today's node

6.2 Convergence

In the next two sections, we examine the performance of the pricing techniques. Our benchmark will be the Jamshidian pricer, since it gives an analytical solution. However, we should note that it is not perfect[2], but as far as we are concerned, there is no better method to give a more accurate price. We are interested in three things: (1) How accurate is the result given by the pricer? (2) How fast is the engine? (3) How sensitive is it to certain parameters of the product or model?

We inspect each of these questions for every model (the HW, the CIR, and the Vasicek) and pricing method (three different Monte Carlo engines and trees). We shall start with measuring accuracy, that is, convergence of the Monte Carlo methods, in this section, and we shall continue with parameter sensitivity in the next section.

Our baseline scenario consists of a swap, a yield curve and a base model of each kind. The exact parameters are detailed in Table 6.2. n refers to the number of time steps in each simulation and M is the number of simulations. The yield curve is that of Hull & White (1994*b*) to make results of the tree comparable to the original paper. Of course, the baseline scenario contains redundant parameters: for the Hull–White model, we do not need ϑ or r_0, but we have a yield curve and the opposite is true for the other two models.

$\varkappa =$	0.1	tenor:	3	coupon:	0.02
$\vartheta =$	0.08	frequency:	annual	increment:	0
$\sigma =$	0.01	expiry:	2 years	$n =$	100
$r_0 =$	0.03	yield curve:	$0.08 - 0.05e^{-0.18t}$	$M =$	50 000

Table 6.2: Baseline scenario for simulations.

First, we run a Milstein, a Linear Drift, and an Euler–Maruyama Monte Carlo with $M = 100\ 000$ for each of the models, and measure their accuracy and run time. The running mean is presented in Figure 6.2, whereas signed relative error terms are shown in Figure 6.3. Table 6.3 details run time and the standard deviation of pricing errors across the 100 000 simulations.

It is apparent that more or less all methods in all models converge to the true swaption

[2]See e.g. numerical instability in the CIR model in Section 6.3. When coupon levels are high, the Jamshidian pricer gives a small, but negative value, which is, by definition, impossible. It can be corrected if we replace all negative values by 0s, but our aim is to point out any limitations to the application of the methods, and it serves this purpose better if we keep errors as they are.

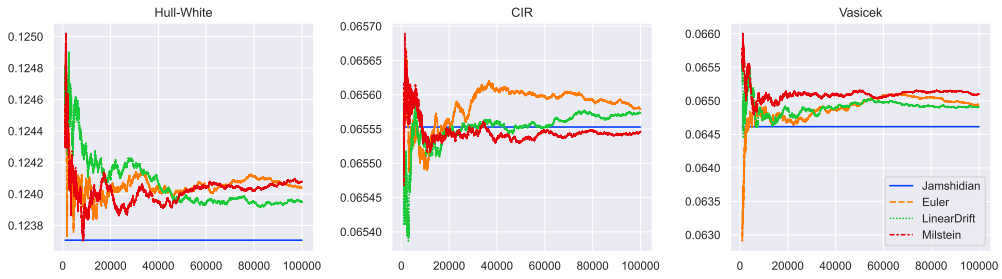


Figure 6.2: Convergence of Monte Carlo methods in different models.

value (or somewhere very close). The speed of convergence, however, differs between models. In the CIR case, it is fast, and we get reasonable results in a few thousand runs. At 10 000, the signed relative pricing error is below 0.1% for all models. At 100 000, it is very close to zero with little deviation. In the CIR model, the Milstein scheme is the first to converge and the Euler has the highest variance. The run time is also really good, 9-36 seconds, but note that the Milstein is about 3-4 times longer. This is because in this model both second derivatives exist and the Milstein scheme consists of four extra factors and two sets of random Gaussian variables.

Figure 6.3: Signed relative error of MC methods as function of the number of paths.

Compared to the CIR, the Hull–White is slower in convergence, at 10 000 its error terms are below 0.5% and at 100 000 they seem to converge to 0.3%. Note that while in the CIR case, error terms, especially the Linear Drift one varied between positive and negative values, in the Hull–White model, they are mostly positive. It is more difficult to say which method starts to converge first, but at 100 000 the Linear Drift seems the best option. An interesting notion is that run times are about 25 times higher for the first-order schemes, but the increase is relatively small in the Milstein case, only 12-13 seconds. This can be explained by the no-arbitrage nature of the model. While both the CIR and the Vasicek have endogenous bond prices, the Hull–White pretty much depends on the yield curve

and requires array searches for each bond pricing. It has higher standard errors than the CIR as well.

	Vasicek		Hull–White		CIR	
	time (s)	std error (%)	time	std error	time	std error
Euler	**7**	**43.10**	**245**	20.85	**9**	8.12
Linear Drift	8	43.32	246	20.82	10	8.05
Milstein	18	43.34	258	**20.81**	36	**8.04**

Table 6.3: Performance of Monte Carlo models.

Finally, the Vasicek takes the least time to run, but comes with a price of high standard deviation. This is a peculiarity of the model: while volatility is mitigated by the square root term (r can be assumed to be below 100%) in the CIR model, it remains high in the Vasicek case. It seems to be a mixture of the previous models: on the one hand at first it is not uncommon to see negative pricing errors, but on the other hand, convergence is rather slow, and at 100 000 runs it does not seem to move much away from 1%. Fortunately, it has similar magnitudes at 50 000, so in the following sections we set $M = 50\ 000$. To conclude, it is hard to say which model is best suited to either the Hull–White or Vasicek model; all three have roughly equal performance.

6.3 Parameter Sensitivity

Next, we inspect the parameter sensitivity for each of the models and methods. The scenarios are shown in Table 6.4. Notice that parameter arrays are different across pricing methods. It is motivated by the fact that the Hull–White model is a lot slower, so we reduce the number of cases to mitigate that. Furthermore, some models are highly sensitive to e.g. mean reversion speed. In particular, the Hull–White tree, in its current form, throws an error outside this narrow range.

Results are plotted in Figures 6.4, 6.6, 6.5 for Hull–White, Vasicek and CIR Monte Carlos, respectively, while Figures 6.7, 6.9 and 6.8 report those of the trees. Figure 6.11 presents sensitivity to number of steps in the CIR and Vasicek MC. The signed relative error of pricing for each method and parameter can be found in Appendix A.

	HW	V & CIR	V & CIR tree	HW tree
$\varkappa$	0.01, 0.05, 0.1, 0.2	0.01 to 0.295	0.01 to 0.2	0.095 to 0.105
ϑ	-	0.02 to 0.21	0.02 to 0.11	-
σ	0.01, 0.02, 0.05, 0.1	0.01 to 0.06	0.01 to 0.06	
r_0	-	0.01 to 0.2	0.01 to 0.1	-
tenor	2, 5, 10, 20	2 to 40	2 to 20	
increment	-0.05 -0.01, 0, 0.01, 0.05	-0.05 to 0.05	-0.05 to 0.05	
freq	monthly, quarterly, semi-annual, annual			
expiry	1, 2, 5, 10	1 to 10	1 to 10	
coupon	0.01, 0.02, 0.04, 0.06	0.01 to 0.1	0.01 to 0.1	
yield curve	Table 6.5	-	-	Table 6.5
n	5 to 1000	5 to 1000	5 to 1000	

Table 6.4: Sensitivity analysis scenarios.

6.3.1 Mean reversion Speed

First, we inspect how the swaption price changes with respect to the mean reversion speed. It was denoted by $\varkappa$ in all models.

In the Hull–White Monte Carlo, prices change very little in the range $[0.01, 0.2]$, with error terms varying between $0.2-0.4\%$. Mean reversion speed on the higher end generally results in more accurate pricing, but only the Milstein is monotone. For lower values, it produces a smaller error, but from 0.1, the Euler method performs better. As mentioned, the tree is highly sensitive to this value, because the construction of the grid depends heavily on $\varkappa\Delta$, and we fixed Δ to match the frequency, so it is quite large. In the narrow range, though, where the tree is able to price, it gives a consistent, slightly overpriced value.

Contrary to the Hull–White model, in the Vasicek and CIR models prices *do* vary a lot with mean reversion speed. Error terms in the CIR model are smaller in magnitude with a maximum of 0.15%, and they are oscillating to 0, while in the Vasicek model they are much higher with a maximum of 1.5% reducing slowly to 0. In general, higher mean reversion speed means smaller error, and the swaption price is a concave increasing function of $\varkappa$ in the given interval, ranging from 0.04 to 0.1. The trees perform much worse in these cases, with errors up to -5% in the CIR and 2% in the Vasicek model.

6.3.2 Long-Term Mean

The long-term mean was denoted by ϑ and is only present in the CIR and Vasicek models.

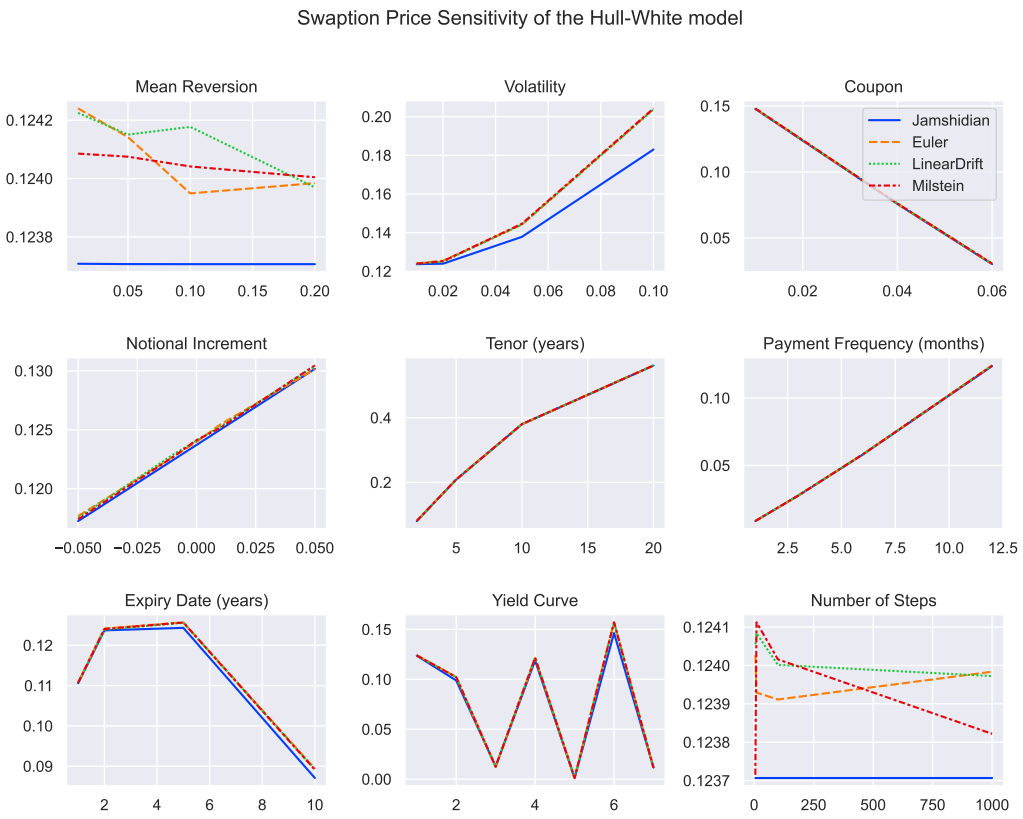


Figure 6.4: Swaption price sensitivity to parameters in the Hull–White Monte Carlo.

In both models, the price increases about $3-5$-fold in the given range, and seems close to linear. All methods perform roughly the same for higher values, and the Linear Drift has the smallest variance. The error terms are between -0.1% and 0.2% for the CIR model and decrease from 1% to close to 0 in the Vasicek. The trees again give much worse results with errors starting from 2 or 10%. Pricing is most accurate when the long-term mean is close to the initial short rate value r_0.

6.3.3 Short Volatility

Short volatility is a somewhat confusing term, and what we mean is the parameter σ in all models, which is of lesser importance in the CIR model, as it is only *part* of the volatility term.

It is fair to say that the price reacts the most visibly to this parameter, with Monte Carlo prices largely deviating from the theoretical value. But, while this deviation is positive

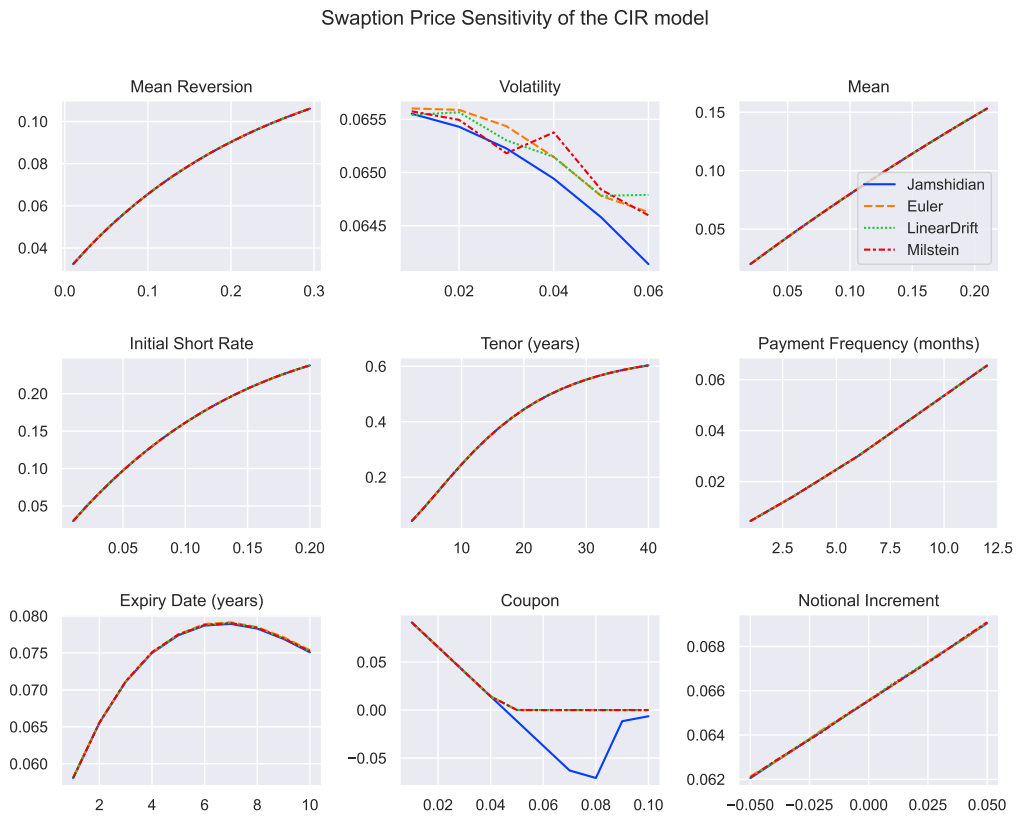


Figure 6.5: Swaption price sensitivity to parameters in the CIR Monte Carlo.

for all models and is highest for the Hull–White with errors reaching 10% at $\sigma = 0.1$, the shape of the curve is very different across models. Peculiarly, in the CIR case, the curve is downward-sloping, while in the other two models, it is a slightly increasing convex function and varies the most in the HW case. The performance of methods is roughly equal, it is hard to pick a champion in either model.

6.3.4 Tenor

The price sensitivity to tenor looks the same for all models: an upward-sloping concave curve. The CIR and Vasicek trees are, however, highly sensitive to time variables, and produce errors of 10^{30} magnitude for years more than 5, so they had to be excluded from the figures. In Monte Carlo-s, errors are about the same magnitude as for mean reversion speed, fading away at the higher end. It is unbeknownst to us why longer-dated products are priced more accurately, but the cause can perhaps be searched in the mean reverting

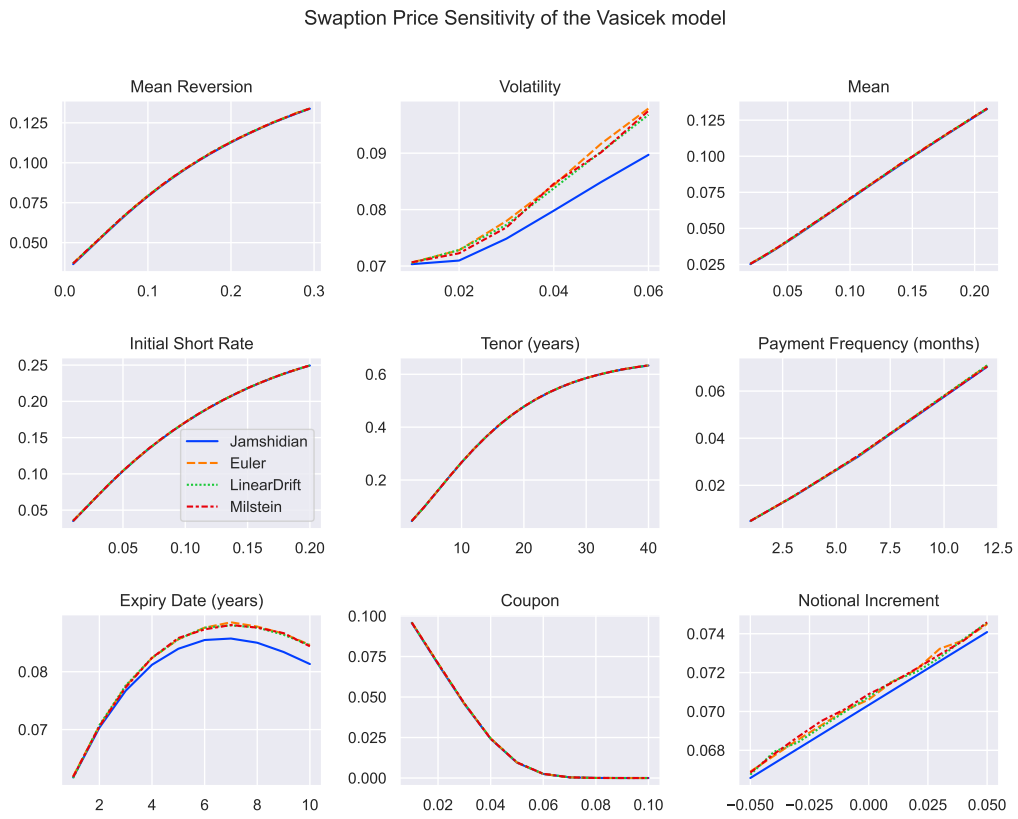


Figure 6.6: Swaption price sensitivity to parameters in the Vasicek Monte Carlo.

property: variance reduces over time in all these models. There is no obvious winner: for the HW, it is the Linear Drift model; in the CIR it is the Milstein, while in the Vasicek, the Euler.

6.3.5 Notional Increment

The notional increment would be more relevant in Bermudan cases, where accreters and amortisers are much common than in European settings.

As can be expected, accreters are priced higher and roughly all methods perform the same.

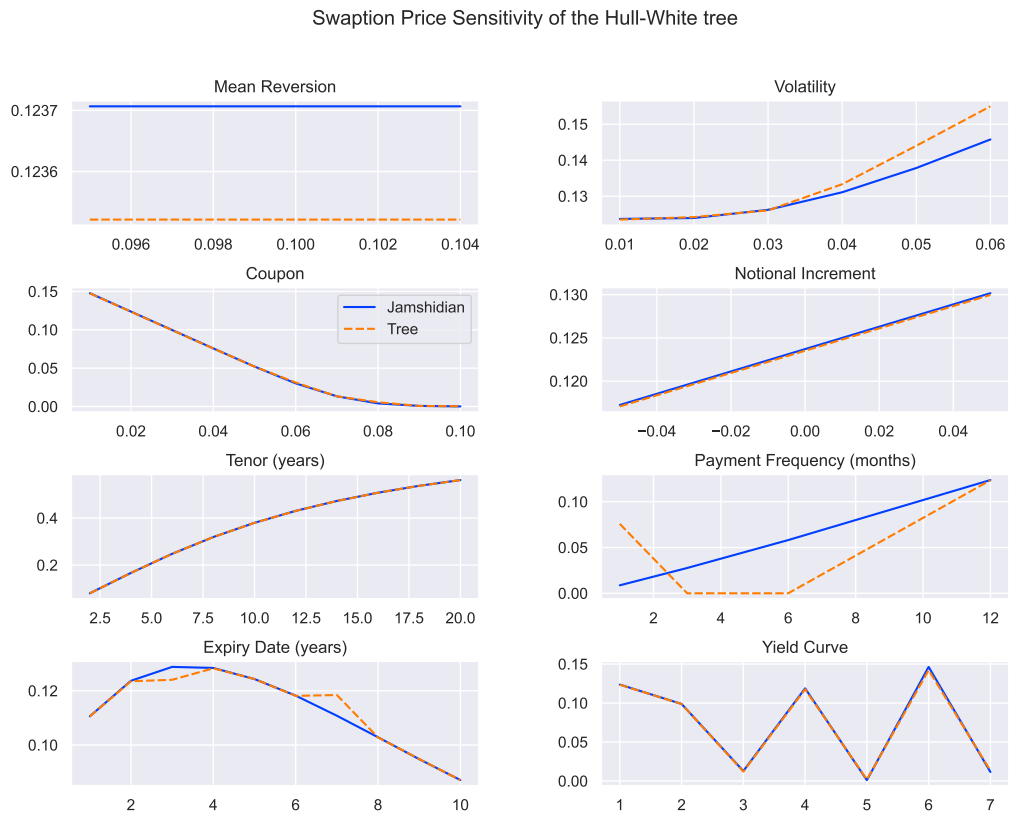


Figure 6.7: Swaption price sensitivity to parameters in the Hull–White tree.

6.3.6 Payment Frequency

The payment frequency is strongly related to the tenor of the swap and exhibits the same characteristics.

6.3.7 Expiry Date

The swaption price is not monotone for either of the models in the time to expiry. First, the curve slopes upward until it reaches its peak at about 6-8 years, then starts to decrease. It is in line with facts from the Black & Scholes (1973) world. Curiously enough, pricing is now less accurate at the far end, which somewhat contradicts the results in the tenor section. The trees become numerically instable, as with all time variables.

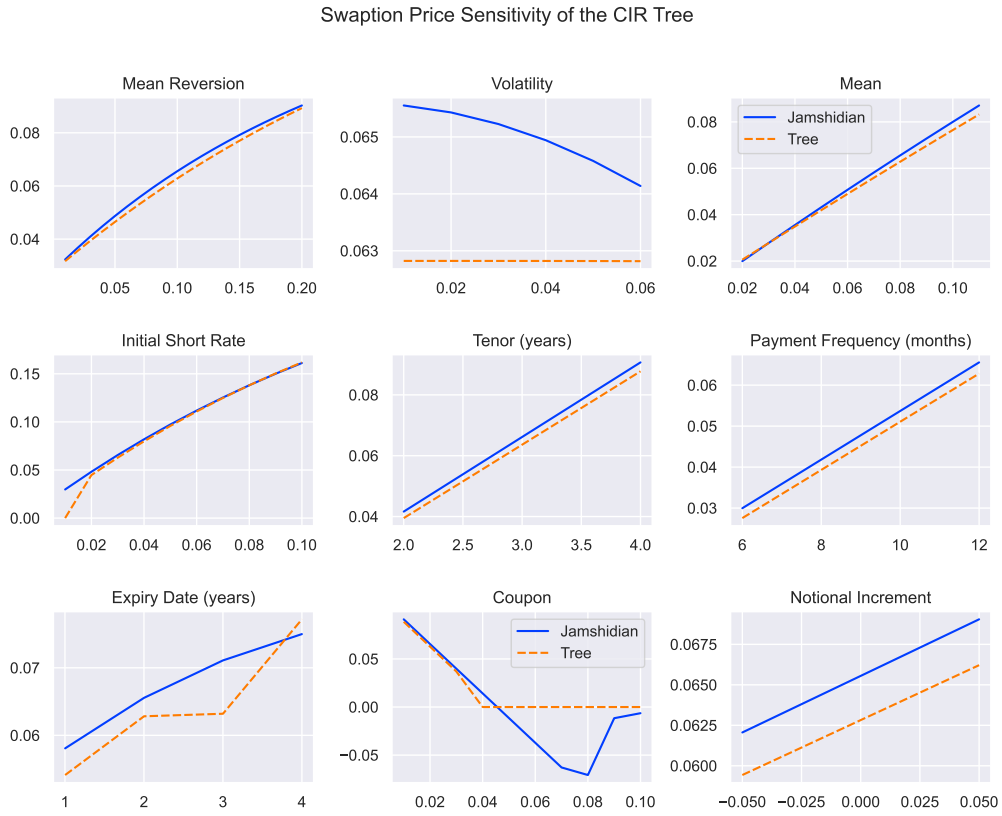


Figure 6.8: Swaption price sensitivity to parameters in the CIR tree.

6.3.8 Coupon

The payer swaption can be viewed as a call option on the swap rate, struck at the fixed coupon. It is thus natural that we see a downward-sloping convex function, which levels off at zero around the par rate of the swap. Pricing is usually accurate, especially for lower coupon values, but error terms reach -100% in the Vasicek and CIR case, since the engine often returns exactly zero (meaning that not one simulation out of 50000 has produced a positive price). We shall note that in this instance the Jamshidian pricer itself cannot be fully trusted, since it returns negative values in the CIR case, which contradicts definition of the swaption. The cause must lie in computing non-central chi-squared cumulative distribution function values of different non-centrality parameters.

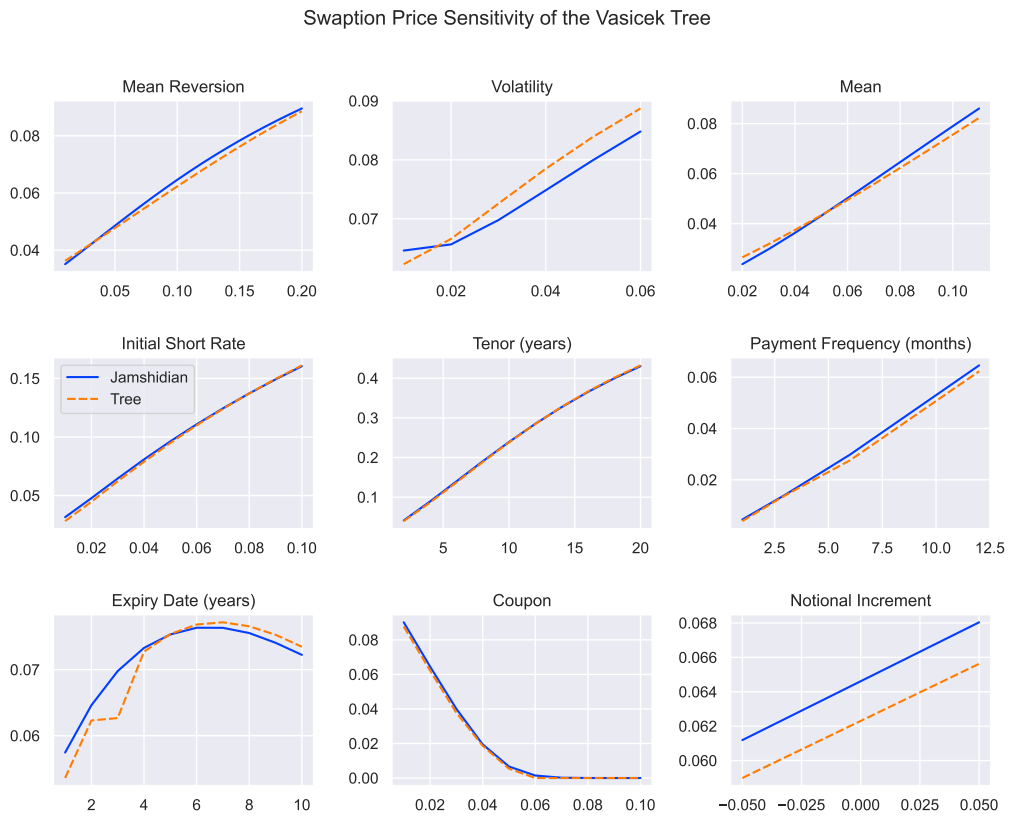


Figure 6.9: Swaption price sensitivity to parameters in the Vasicek tree.

6.3.9 r_0

The initial short rate is not much different from the long-term mean in the two equilibrium models, and exhibits the same characteristics.

6.3.10 Yield Curve

We only need to examine yield curve sensitivity in the no-arbitrage Hull–White model. We propose seven scenarios, some fictional, some based on historical or present data. Table 6.5 and Figure 6.10 detail these.

Of these, Scenarios 1, 2, 4, and 6 are high-yield scenarios, with 6 being the highest (at the very beginning of the 2008–2009 crisis). Scenario 4 (the current Hungarian situation) is very similar to 2, a fictional high flat curve.

The price is always higher in high-yield scenarios. Though it was not obvious beforehand,

1:	$0.08 - 0.05e^{-0.18t}$ (Hull & White (1994*b*))
2:	0.06 flat
3:	0.02 flat
4:	based on Hungarian upward-sloping yield curve on 2022.03.25.
5:	based on European negative yield curve on 2022.03.24.
6:	based on Hungarian humped yield curve on 2008.09.25.
7:	based on Hungarian inverse humped yield curve on 2017.09.25.

Table 6.5: Yield curve scenarios (AKK (2022), ECB (2022), Hull & White (1994*b*)).

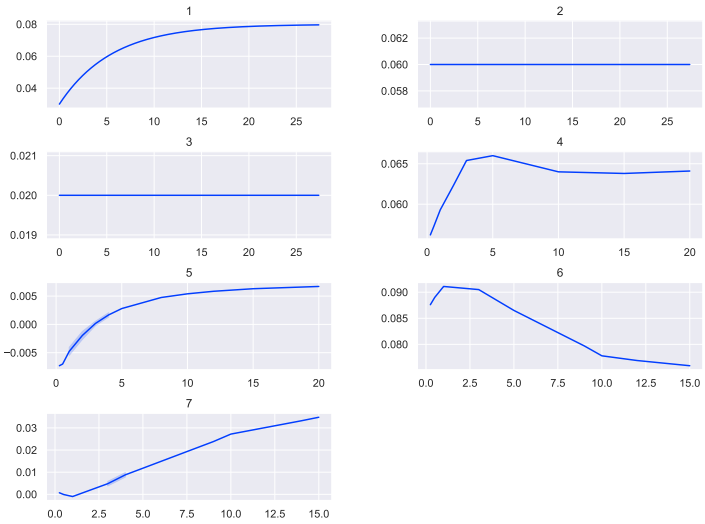


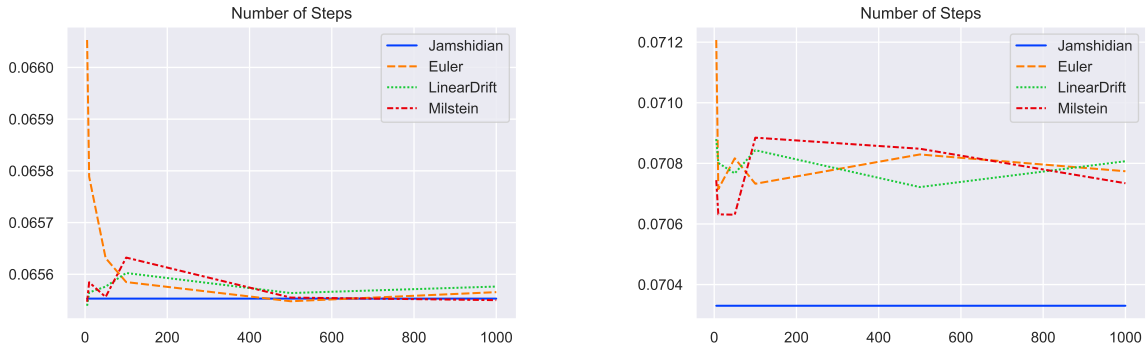
Figure 6.10: Yield curves

it seems that the fixed leg is more sensitive to yield curve movements because both legs contain zero coupon bonds, and the floating leg is decreasing in rates and the fixed (negative) leg is increasing. Pricing errors are highest for the 5th, current European negative yield scenario, and lowest for the 3rd, low flat curve. The same is true for the tree pricer, but the error term in the crisis becomes really high: 60%.

6.3.11 Number of Steps

Finally, we inspect the number of steps in the Monte Carlo simulation. In the Hull–White model, convergence is slow, and the Milstein scheme is the obvious champion, but at 1000 steps it still has about 0.1% of pricing error. In the CIR model, convergence is fast for all models, and at 500 steps, the Milstein prices almost exactly as the Jamshidian pricer.

The Vasicek is again the least accurate, it does not seem to converge at all, the three methods seem to level off at 0.6% pricing error.



(a) CIR (b) Vasicek

Figure 6.11: Swaption price sensitivity to the number of steps in Monte Carlo.

6.4 Further Improvements

After implementing three different type of pricers for three models, and examining performance and parameter sensitivity of these methods, we propose ways to further improve the analysis presented above:

1. One major drawback of the implementation of trees is that currently the time step must equal the payment frequency, which, when talking about annual swaps, is a significantly large quantity, the whole tree heavily relying on it. A very straightforward improvement would be to reduce this Δ. This would require a substantially larger tree with a lot of pricing steps, but while this would only lead to a slightly increased computational time, we should also solve, and this is a more complex problem, that at *not all* steps should the zero coupon bond prices be calculated, only at dates corresponding to payments.

2. Currently the binomial trees are quite simple, a more elaborate tree design could be more appropriate, but as equilibrium models are out of interest in industry, this shortcoming is negligible.

3. We presented three swaption pricing techniques: an analytical approach and two numerical schemes. We omitted, however, two of the most popular numerical meth-

ods: finite differences and approximations. Further research could investigate in which cases they are more efficient and preferable to Monte Carlo-s or trees.

4. Though briefly mentioned, we have not implemented the implicit Euler–Maruyama discretisation scheme, nor have we employed variance reduction techniques such as antithetic sampling. Both measures can improve precision of the Monte Carlo.

5. We only covered almost-vanilla swaptions, with only the notional having a proper term structure. A possible development would be to extend the methods to swaptions with time-varying coupons and spreads.

6. Although European swaptions are heavily traded on markets, perhaps Bermudans are even more popular. Another possible improvement of the techniques presented above could be an extension to path-dependent instruments. We suspect that trees could be used without much alteration, but contrary to that, we should fundamentally change the way how the Monte Carlo engines are built. Furthermore, there is only a slim chance that analytical formulae exist for non-European options.

7. We presented the Vasicek, the CIR and the Hull–White model. While the former two are equilibrium models and thus have very little to do with actual market data, we could have calibrated the Hull–White against market-observed swaption prices. One problem is, nonetheless, that the one-factor Hull–White model still only has two free parameters and hence cannot possibly price more than two instruments precisely. A straightforward extension is to instead examine the general one-factor Gaussian model, e.g. with piece-wise flat mean reversion speed and volatility, and it is indeed used in industry. Nevertheless, this model is still not able to correctly capture, for instance, volatility smile or imperfect correlation of assets.

8. Elaborating on the previous remark, our model choices, specifically the one-factor property, limited the cases to which they are applicable. A multi-factor extension, e.g. of the Gaussian model, is a far more acceptable option, and Andersen & Piterbarg (2010) even derives an analytical, Jamshidian-based approach for the multi-factor Hull–White.

9. Finally, we only examined models where the drift term is linear and the diffusion term is a deterministic function of the short rate. The most up-to-date models in industry, however, are stochastic volatility models. Further research thus should focus on how swaption pricing is done in either the SABR or the Cheyette model which are just two of the most state-of-the-art frameworks in interest rate modelling.

Chapter 7

Conclusion

Swaptions are one of the most liquid products in interest rate markets and are often the calibration tool of models. Therefore, it is paramount that their pricing is as accurate as possible. In this dissertation, we have presented three swaption pricing techniques in different model settings: the analytical Jamshidian decomposition and the numerical Monte Carlo and tree methods.

Our analysis showed that when the decomposition exists, it works well, except for a few cases in the CIR model where slight numerical errors were produced. The problem with the method is its (non-)existence. We have built numerical engines and compared them to the Jamshidian pricer to investigate which could be a good replacement if analytical solutions do not work.

Performance and convergence tests identified that a Monte Carlo engine can be a good alternative in the CIR model: it is quick (takes about 10-20 seconds to run 100 000 simulations), accurate and not too sensitive to parameter shocks.

In the Vasicek model, Monte Carlo methods are also very fast, but, due to the model specification, pricing errors are of greater magnitude and convergence is slower. However, it is still better than the Nelson & Ramaswamy (1990) binomial tree approach, which is less precise, only slightly faster, and highly sensitive to time parameters.

In the Hull–White framework, the Monte Carlo is about as accurate as the trinomial tree, but takes significantly more time to produce a result. It is 25 times slower than in the other two models, and its variance is higher than in the CIR case. The tree, on the other hand, albeit quick, is super sensitive to mean reversion speed, which is a fundamental building block of the tree structure. By improving the tree with varying and smaller time-step sizes, this shortcoming could be mitigated.

As for different Monte Carlo discretisation schemes, there is no obvious champion model. Each has its merits and performs best in certain scenarios. Nevertheless, we would propose that the Milstein scheme be used in the CIR case and the Linear Drift Euler–Maruyama scheme in the other two models.

To conclude, there is no superior method and one should carefully consider all important aspects before deciding on the pricing technique. One thing is certain, though: *if* an analytical Jamshidian decomposition exists, it is both very fast and close to the results of Monte Carlo and tree schemes, so it is fair to say that it indeed returns the true no-arbitrage price of the asset.

Glossary

accreter növekvő névértékű swap. 8

accrual period kamatfelhalmozódási időszak. 6

affine model affin modell. 39

affine short rate models affin rövidkamatláb-modellek. 58

amortiser amortizáló swap. 8

annuity annuitás. 6

Bermudan swaption Bermuda swaption. 12

binomial tree binomiális fa. 16

bond option kötvényre szóló opció. 30

call option call opció; vételi jog/eladási kötelezettség. 11

continuously compounded forward yield folytonos forward ráta. 4

convergence konvergencia. 20

coupon kupon. 7

coupon-bearing bond kamatozó kötvény. 32

cross-currency swap cross-currency swap; kamatcsereügylet, melynek két lába eltérő devizában denominált. 8

deposit betét. 3

diffusion term diffúziós tag. 20

drift drift. 20

equilibrium model egyensúlyi modell. 29

European swaption európai swaption. 12

exercise date lehívás dátuma. 11

exercise price kötési ár. 11

finite difference method véges differenciák módszere. 16

fixed leg fix láb. 7

floating leg változó láb. 7

geometric Brownian motion geometriai Brown mozgás. 34

grid rács, háló. 16

instantaneous forward rate pillanatnyi, folytonos forward ráta. 5

interest rate derivative kamatláb-derivatíva. 7

interest rate model kamatlábmodell. 28

Law of Large Numbers Nagy Számok Törvénye. 20

Markovian Markov tulajdonságú. 48

martingale martingál. 11

maturity lejárat. 3

mean reversion átlaghoz visszahúzás. 30

moment matching momentumok módszere. 17

multi-factor model többfaktoros modell. 18

no-arbitrage model no-arbitrage modell. 28

notional névérték. 8

numéraire ármérce. 11

one-factor model egyfaktoros modell. 32

option opció. 11

Ornstein–Uhlenbeck process Ornstein–Uhlenbeck folyamat. 35

par rate par vagy névleges kamatláb. 6

path-dependent instrument útfüggő derivatíva. 16

pay-off function kifizetésfüggvény. 11

payer swap payer swap; csereügylet, melyben a fix lábat fizetjük. 7

put option put opció; eladási jog/vételi kötelezettség. 11

receiver swap receiver swap; csereügylet, melyben a változó lábat fizetjük. 7

recombining tree összeölelkező fa. 17

reference rate referenciahozam. 8

risk-free rate of return kockázatmentes hozam. 17

risk-neutral measure kockázatsemleges mérték. 11

short rate azonnali vagy rövid kamat. 5

simple forward rate egyszerű vagy LIBOR forward ráta. 4

spot rate azonnali vagy rövid kamat. 5

spread spread, felár. 8

stability stabilitás. 20

stochastic differential equation sztochasztikus differenciálegyenlet. 33

stochastic volatility model sztochasztikus volatilitás modell. 32

strike kötési ár. 11

swap csereügylet, itt: kamatcsereügylet. 7

swap rate swap ráta. 6

swaption swaption, csereügyletre szóló opció. 11

tenor futamidő. 4

Term Structure Equation Lejárati Szerkezet Egyenlet. 38

tree method fa módszer. 15

trinomial tree trinomiális fa. 18

underlying alaptermék. 11

vanilla swap vanília swap: időben állandó névérték, kupon és spread. 8

volatility surface volatilitásfelület. 30

yield hozam. 3

yield curve hozamgörbe. 12

zero coupon bond elemi kötvény, zéró kupon kötvény. 3

Bibliography

AKK (2022), 'AKK Zero coupon bond yields'.
URL: *https://akk.hu/statisztika/hozamok-indexek-forgalmi-adatok/zero-kupon-hozamok*

Andersen, L. B. & Piterbarg, V. V. (2010), *Interest Rate Modeling*, Atlantic Financial Press.

Antonov, A., Konikov, M. & Spector, M. (2015), 'The free boundary SABR: natural extension to negative rates', *Available at SSRN 2557046* .

Baldi, P. (2017), *Stochastic calculus*, Springer.

Black, F., Derman, E. & Toy, W. (1990), 'A one-factor model of interest rates and its application to treasury bond options', *Financial analysts journal* **46**(1), 33–39.

Black, F. & Karasinski, P. (1991), 'Bond and option pricing when short rates are lognormal', *Financial Analysts Journal* **47**(4), 52–59.

Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**(3), 637–654.

Brigo, D. & El-Bachir, N. (2010), 'An exact formula for default swaptions' pricing in the SSRJD stochastic intensity model', *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics* **20**(3), 365–382.

Brigo, D. & Mercurio, F. (2001), 'A deterministic–shift extension of analytically–tractable and time–homogeneous short–rate models', *Finance and Stochastics* **5**(3), 369–387.

Brigo, D. & Mercurio, F. (2006), *Interest rate models-theory and practice: with smile, inflation and credit*, Springer.

Chen, R.-R. & Scott, L. (1992), 'Pricing interest rate options in a two-factor Cox–Ingersoll–Ross model of the term structure', *The review of financial studies* **5**(4), 613–636.

Cheyette, O. (2001), 'Markov representation of the Heath–Jarrow–Morton model', *Available at SSRN 6073* .

Choi, J. & Shin, S. (2016), 'Fast swaption pricing in Gaussian term structure models', *Mathematical Finance* **26**(4), 962–982.

Collin-Dufresne, P. & Goldstein, R. S. (2002), 'Pricing swaptions within an affine framework', *The Journal of Derivatives* **10**(1), 9–26.

Cox, J. C., Ingersoll Jr, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**(2), 385–408.

Cox, J. C., Ross, S. A. & Rubinstein, M. (1979), 'Option pricing: A simplified approach', *Journal of financial Economics* **7**(3), 229–263.

Crank, J. & Nicolson, P. (1947), A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type, *in* 'Mathematical Proceedings of the Cambridge Philosophical Society', Vol. 43, Cambridge University Press, pp. 50–67.

Di Francesco, M. & Kamm, K. (2022), 'On the deterministic-shift extended CIR model in a negative interest rate framework', *arXiv preprint arXiv:2203.07458* .

Dothan, L. U. (1978), 'On the term structure of interest rates', *Journal of Financial Economics* **6**(1), 59–69.

Duffie, D. & Kan, R. (1996), 'A yield-factor model of interest rates', *Mathematical finance* **6**(4), 379–406.

Dybvig, P. H. (1997), 'Bond and bond option pricing', *Mathematics of Derivative Securities* **15**, 271.

ECB (2022), 'Yield curves - ECB Statistical Data Warehouse'.
URL: *https://sdw.ecb.europa.eu/browse.do?node=9691126*

Feller, W. (1951), 'Two singular diffusion problems', *Annals of mathematics* pp. 173–182.

Gatarek, D. & Jabłecki, J. (2021), 'Between Scylla and Charybdis: The Bermudan swaptions pricing Odyssey', *Mathematics* **9**(2), 112.

Glasserman, P. (2004), *Monte Carlo methods in financial engineering*, Vol. 53, Springer.

Hagan, P. S. (2000), Methodology for callable swaps and Bermudan "exercise into" swaptions, Technical report, Discussion Paper.

Hagan, P. S., Kumar, D., Lesniewski, A. S. & Woodward, D. E. (2002), 'Managing smile risk', *The Best of Wilmott* **1**, 249–296.

Heath, D., Jarrow, R. & Morton, A. (1992), 'Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation', *Econometrica: Journal of the Econometric Society* pp. 77–105.

Henrard, M. (2003), 'Explicit bond option formula in Heath–Jarrow–Morton one factor model', *International Journal of Theoretical and Applied Finance* **6**(01), 57–72.

Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *The review of financial studies* **6**(2), 327–343.

Ho, T. S. & Lee, S.-B. (1986), 'Term structure movements and pricing interest rate contingent claims', *the Journal of Finance* **41**(5), 1011–1029.

Hull, J. (1996), 'Using Hull–White interest rate trees', *Journal of Derivatives* **3**(3), 26–36.

Hull, J. C. & White, A. D. (1994*a*), 'Numerical procedures for implementing term structure models ii: Two-factor models', *The Journal of Derivatives* **2**(2), 37–48.

Hull, J. & White, A. (1990), 'Pricing interest-rate-derivative securities', *The review of financial studies* **3**(4), 573–592.

Hull, J. & White, A. (1993), 'One-factor interest-rate models and the valuation of interest-rate derivative securities', *Journal of financial and quantitative analysis* **28**(2), 235–254.

Hull, J. & White, A. (1994*b*), 'Numerical procedures for implementing term structure models i: Single-factor models', *Journal of derivatives* **2**(1), 7–16.

Jamshidian, F. (1989), 'An exact bond option formula', *The journal of Finance* **44**(1), 205–209.

Jamshidian, F. (1995), 'A simple class of square-root interest-rate models', *Applied Mathematical Finance* **2**(1), 61–72.

Karatzas, I. & Shreve, S. (2012), *Brownian motion and stochastic calculus*, Vol. 113, Springer Science & Business Media.

Kloeden, P. E. & Platen, E. (1992), *Numerical solution of stochastic differential equations*, Springer Berlin Heidelberg.

Le Gall, J.-F. et al. (2016), *Brownian motion, martingales, and stochastic calculus*, Vol. 274, Springer.

Leclerc, M., Liang, Q. & Schneider, I. (2009), 'Fast Monte Carlo Bermudan Greeks', *Risk* **22**(7), 84.

Leippold, M. & Wiener, Z. (2003), 'On trinomial trees for one-factor short rate models'.

Longstaff, F. A. & Schwartz, E. S. (1992), 'Interest rate volatility and the term structure: A two-factor general equilibrium model', *The Journal of Finance* **47**(4), 1259–1282.

Maghsoodi, Y. (1996), 'Solution of the extended CIR term structure and bond option valuation', *Mathematical finance* **6**(1), 89–109.

Mamon, R. S. (2004), 'Three ways to solve for bond prices in the Vasicek model', *Advances in Decision Sciences* **8**(1), 1–14.

Medvegyev, P. (2002), *Valószínűségszámítás: fejezetek a matematikai analízisből és a valószínűségszámításból*, Aula.

Medvegyev, P. (2014), *Pénzügyi matematika*, Typotex.

Merton, R. C. (1973), 'Theory of rational option pricing', *The Bell Journal of economics and management science* pp. 141–183.

Milshtein, G. (1979), 'A method of second-order accuracy integration of stochastic differential equations', *Theory of Probability & Its Applications* **23**(2), 396–401.

Nelson, D. B. & Ramaswamy, K. (1990), 'Simple binomial processes as diffusion approximations in financial models', *The review of financial studies* **3**(3), 393–430.

Pascucci, A. (2011), *PDE and martingale methods in option pricing*, Springer Science & Business Media.

Rogers, L. C. (2002), 'Monte Carlo valuation of American options', *Mathematical Finance* **12**(3), 271–286.

Schrager, D. F. & Pelsser, A. A. (2006), 'Pricing swaptions and coupon bond options in affine term structure models', *Mathematical Finance* **16**(4), 673–694.

Vasicek, O. (1977), 'An equilibrium characterization of the term structure', *Journal of financial economics* **5**(2), 177–188.

Wei, J. Z. (1997), 'A simple approach to bond option pricing', *Journal of Futures Markets: Futures, Options, and Other Derivative Products* **17**(2), 131–160.

Appendix A

Signed Relative Errors of Pricing Techniques

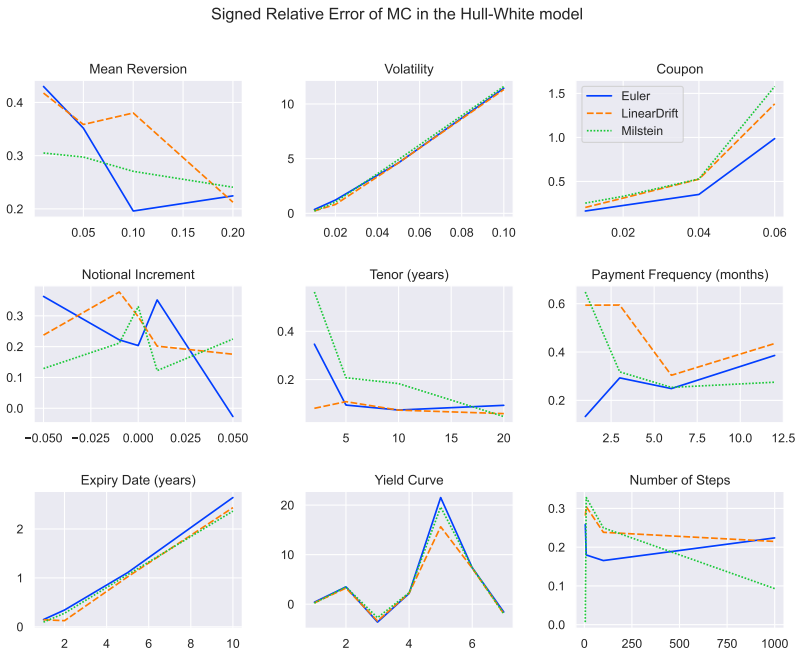


Figure A.1: Signed relative error in the HW MC.

Figure A.2: Signed relative error in the CIR MC.

Figure A.3: Signed relative error in the Vasicek MC.

Figure A.4: Signed relative error in the HW tree.

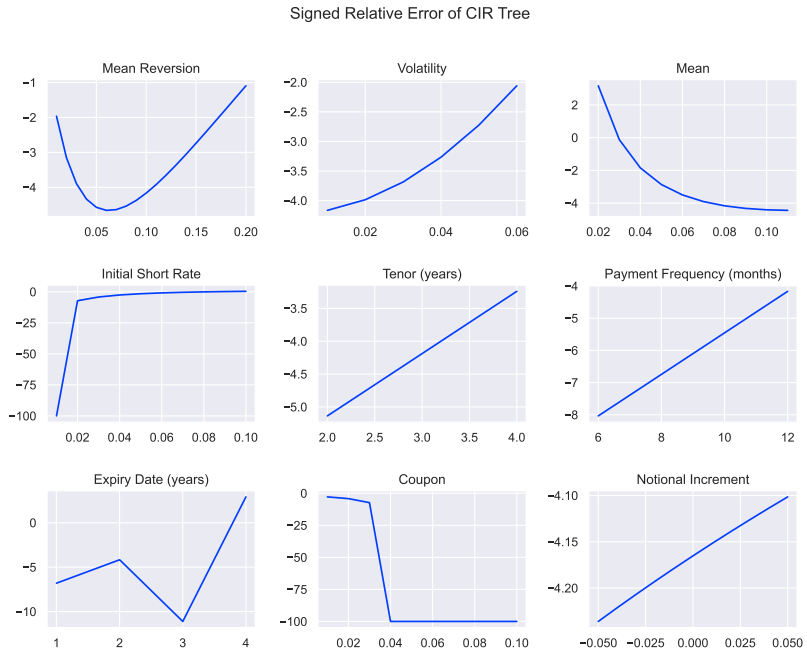


Figure A.5: Signed relative error in the CIR tree.

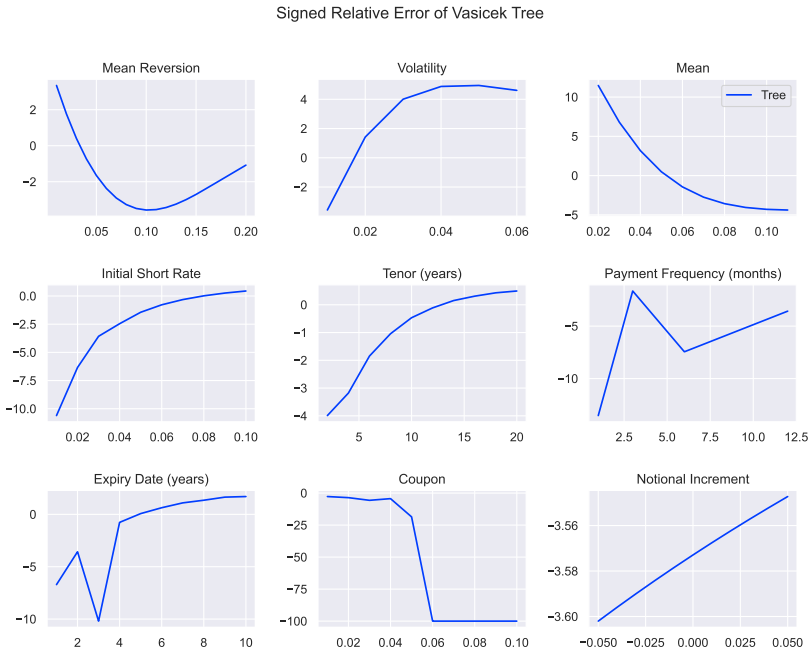
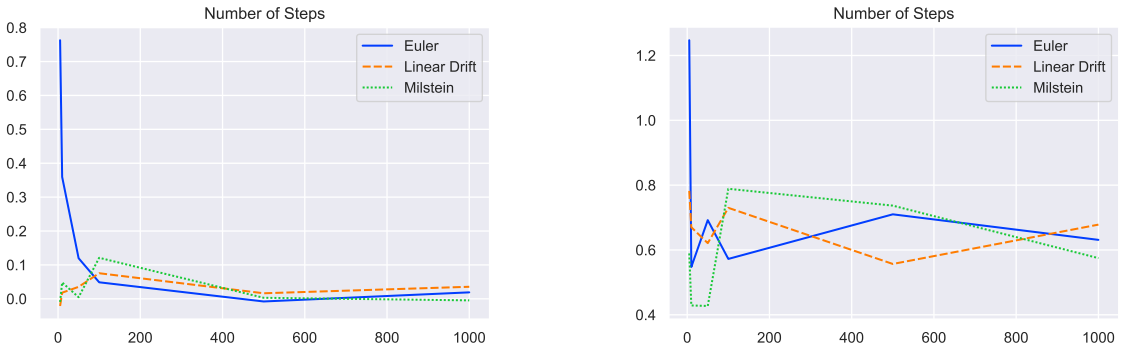


Figure A.6: Signed relative error in the Vasicek tree.



(a) CIR (b) Vasicek

Figure A.7: Signed relative error of MC as function of number of steps.