Control variates for callable Libor exotics

J. Buitelaar

August 2006

Abstract

In this thesis we investigate the use of control variates for the pricing of callable Libor exotics in the Libor Market Model. We introduce the concepts necessary to value these products: interest rate derivatives, the Libor Market Model, Monte Carlo simulation, callable Libor exotics and estimation of the optimal exercise strategy.

For the Bermudan payer (receiver) swaption we show that the cap (floor) is a very good control variate. The reason is that the payoff of a cap and Bermudan swaption are very similar, but it turns out that the results are strongly influenced by the shape of the term structure. We propose methods to improve the variance reduction by looking at other cap-like control variates and find that taking linear combination of caps with different strikes and cash flow dates leads to significant improvements. Finally we show that the results for the Bermudan swaption can be extended to other callable Libor exotics, by taking the capped payoff of the underlying Libor exotic as control variate. For the Bermudan swaption and callable inverse floater we obtain variance reduction factors of order 100. For a snowball, which is path-dependent and has no analytical underlying we obtain a factor 20 reduction in variance.

Contact details	
address	Hugo de Grootstraat 22
	2613TV Delft, Netherlands
tel.	+31 641 211 034
e-mail	Jacob.Buitelaar@gmail.com

Preface

This thesis was submitted in the partial fulfillment for the requirements for the Master's degree in Applied Mathematics at Delft University of Technology. The research for this thesis has been done at the Modelling & Research department of Rabobank International in Utrecht, the Netherlands.

I would like to express my gratitude to the following persons, who have been a great help to me during my work. First of all Roger Lord, who has supervised my research at Rabobank. After helping me to get started, he was always ready to answer my questions, brainstorm with me about new ideas and gave useful feedback on my work. I would also like to thank my other colleagues at Rabobank for their help. Furthermore I would like to thank Hans van der Weide for supervising my thesis at Delft University of Technology, as well as the other members of my graduation committee.

The abstract of this thesis has been submitted for presentation at the 5th Actuarial and Financial Mathematics Day on February 9th, 2007 in Brussels.

Utrecht, August 2006

Jacob Buitelaar

Contents

List of Symbols

Below we present a list of symbols and abbreviations used in this thesis. Onlyfrequently used (i.e. in more than one chapter) symbols are included.

Abbreviations		
AS	antithetic sampling	p.24
ATM	at-the-money	p.6
bp	basis point: 0.01%	p.53
C(c)LE	Cancellable/Callable Libor Exotic	p.36
CcLE	Cancellable Libor Exotic	p.36
CLE	Callable Libor Exotic	p.36
CV	Control Variate	p.25
FRA	Forward Rate Agreemenent	p.3
ITM	in-the-money	p.6
LE	Libor Exotic	p.36
Libor	London InterBank Offer Rate	p.4
LMM	Libor Market Model	p.11
LS	Longstaff-Schwartz	p.41
MC	Monte Carlo	p.22
OTM	out-of-the-money	p.6
Greek letters		
α_i	tenor: the day count between T_{i-1} and T_i	p.4
se	standard error of Monte Carlo simulation	eq.(5.3) p.23
ζ_i	D-vector of factor volatilties	p.11
κ	relative improvement of variance reduction technique	eq.(6.1) p.24
μ_i	drift of L_i	p.11
ψ	normal probability density function	p.34
ρ_{ij}	correlation between W_i and W_j	p.14
σ_i	volatility of L_i	p.14
$\overline{\sigma}_i$	black-volatility of L_i	eq.(3.4) p.12
σ_n^N	volatility of the swap rate	eq.(3.13) p.15
τ	index of optimal exercise date	eq.(9.8) p.41
ϑ	time-adjusted improvent of variance reduction	eq.(6.2) p.24
Roman letters		
$\mathbf{1}_{\{.\}}$	indicator function	p.62
$A_n^N(t)$	annuity or PVBP	eq.(1.7) p.5
$B(t)$	(risk neutral) bank account	eq.(3.7) p.13
$C(k)$	covariance matrix of $\mathbf{L}$ on $[T_k, T_{k+1}]$	eq.(7.7) p.30
capl_i	value of a caplet with maturity T_i	p.7
CF_i	Cash flow at time T_i	p.32
cov	covariance	p.23
C_i	Coupon payment at time T_i	p.36
D	number of independent factors driving the Libor rates	p.11
$\mathbb{E}_B^i$	Expectation under measure Q^B, computed at time T_i	p.22
$f(t)$	instantaneous forward rate	p.3
$F(t;S,T)$	forward rate for period $[S,T]$	p.3
floorl_i	value of a floorlet with maturity T_i	p.7
$H_m(T_i)$	continuation value of CLE	eq.(9.5) p.40
K	principal	p.3

$\mathbf{L}(t)$	vector of Libors L_i	p.11
$L_i(t)$	forward Libor rate with reset date T_{i-1} and maturity T_i	p.4
L_i	Libor rate $L_i(T_{i-1})$	p.4
M	number of Monte Carlo simulations	p.23
$m(t)$	next tenor date: $\min(i : T_i \geq t)$	p.13
N	last tenor date index	p.4
$p(t,T)$	bond price at time t for maturity T	p.3
$PS_n^N(t)$	Value of a $T_n \times (T_N - T_n)$ payer swap	eq.(1.6) p.5
Q^B	risk neutral measre (with $B(t)$ as numeraire)	p.13
Q^i	equivalent Martingale measre with $p(t,T_i)$ as numeraire	p.12
R	strike of fixed rate (FRA, cap, swap)	p.3
$r(t)$	short rate	p.2
$R(t,T)$	spot rate for maturity T	p.2
$R_n^N(t)$	swap rate	eq.(1.8) p.6
$\mathbf{S}$	vector of state variables	p.41
s_f	sample standard deviation of f	eq.(5.2) p.23
$T-e$	exercise moment of C(c)LE	p.39
T_i	tenor date $0 \leq i \leq N$	p.4
$\widetilde{V}$	Monte Carlo estimate of value V	eq.(5.2) p.23
v_k	value of C(c)LE conditional on exercise on T_k	eq.(9.1) p.39
var	variance	p.23
$\mathbf{W}(t)$	D-vector of independent Wiener processes	p.11
$\mathbf{W}^i(t)$	D-vector of independent Q^i-Wiener processes	p.12
W_i	Wiener process driving L_i	p.14

Introduction

This thesis deals with the valuation of callable Libor exotics. Callable Libor exotics are Bermudan-style derivatives whose value depends on Libor forward rates. Because the value depends on different interest rates, we need a multifactor model to price these products. The Libor Market Model is the most appropriate model for this purpose. The first three parts of this thesis will describe all the elements necessary for the valuation of Callable Libor exotics.

In Part I we will describe the Libor market model. We start with a description of plain-vanilla interest rate products and give an overview of the development of interest rate models in the last decades. After introducing the Libor Market Model we will pay attention to calibration issues.

Because the Libor Market model is a multi-factor model, only Monte Carlo-based methods are available for the pricing of derivatives. This will be discussed in Part II. We will start with an overview of the Monte Carlo method. The main disadvantage of the Monte Carlo method is that it converges relatively slow. Depending on the problem under consideration, several methods are available to reduce the variance of the simulation. We will discuss the most appropriate ones for our purpose, especially control variates. Furthermore we will explain how Monte Carlo can be used to price Libor exotics in the Libor Market Model.

In Part III we introduce the derivatives we want to price: callable Libor exotics. We will give their general characteristics and describe how we can value these products. Very important is the estimation of the exercise strategy of these products. We will explain how this strategy can be estimated by the algorithm of Longstaff and Schwartz (2001). In a recent paper Rasmussen (2005) showed how the use of control variates for the valuation of American-style products. We show how the results of this paper can be applied to the valuation of Callable Libor exotics.

After introducing the Libor Market Model, Monte Carlo simulation and callable Libor exotics, we will apply these methods in Part IV. We will look for an generic way to reduce the standard error of the Monte Carlo simulation for callable Libor exotics by using control variates. We start with the Bermudan swaption and we will compare the use of different control variates for different types of swaptions. We start with the use of simple control variates and based on these results we will try to find improvements to get the 'optimal control variate'. Finally we will try to see whether we can apply the same techniques to other callable Libor exotics, being the callable inverse floater and the cancellable snowball.

To the best of our knowledge, the only research that has been published on the use of control variates for (callable) Libor exotics is the work by Jensen and Svenstrup (2005), who look at control variates for the Bermudan swaption. In this thesis we will go much further by looking at a wider range of callable Libor exotics. Moreover, we also present other control variates with clearly better performance and show that these can be applied effectively to all callable Libor exotics under consideration.

This reader of this thesis is supposed to be familiar with the basics of financial calculus. In appendix A we give an overview of literature that could be consulted to obtain the required knowledge.

Part I
Libor Market Model

1 Interest rate derivatives

Interest rate derivatives are products whose payoffs are dependent on the level of interest rates. Until the 1970's, the interest rate market mainly consisted of bonds. During the last decades, the volume of trading in other interest rate derivatives, over the counter or on an exchange, increased very quickly. This chapter describes some of the most common, simple (also called plain vanilla) interest rate products. These products form the fundaments of interest rate models and are the building blocks of more complicated (exotic) interest rate derivatives, which will be discussed in Chapter 8. More elaborate discussions on the products described in this chapter can be found, for example, in (Hull, 2003).

1.1 Definitions

Suppose we are standing at time t. Then we can define the following interest rates (Björk, 2004):

- $R(t,T)$: the (simply-compounded) spot rate. This is the interest rate we earn on an investment over the period $[t,T]$, where T is called the maturity date. For each fixed T it is a function of time t for $t < T$.
- $r(t)$: the short rate. The instantaneous interest rate we earn at time t. It is defined as $\lim_{T \downarrow t} R(t,T)$.

From the spot rate $R(t,T)$ we can define the term structure. For a given time t it is given by the function $R(t,T)$, for $T > t$. See Figure 1.1 for an example of a term structure. Generally, $R(t,T)$ is only observable from the market for a finite number of

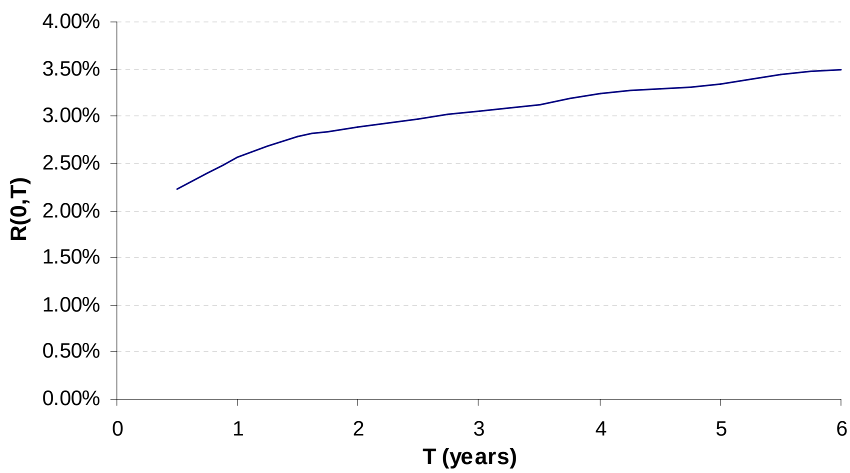


Figure 1.1: Term structure example

maturity dates T. The term structure is estimated by interpolating between these dates. When the term structure is an increasing function of T, it is called upward sloping or normal, which is usually the case. When the term structure is decreasing in T it is called downward sloping or inverted.

Besides these, we also define forward rates. A forward rate is the rate of interest that applies to a future period of time. Suppose we are standing at time t and we would like to invest an amount of money over a future time period $[S, T]$. The interest is paid at the end of the period, at time T. The interest rate we can receive over this period, when contracted at time t is called the forward rate, and denoted by $F(t, S, T)$ for $t < S < T$. Again, T is the maturity date, S is called the reset or settlement date. The value of the forward rate can change over time, as long as $t < S$. At $t = S$, its value is settled and will stay fixed. Similar to the spot rate, we can also define the instantaneous forward rate $f(t, S) \equiv \lim_{T \downarrow S} F(t; S, T)$. Furthermore note that $R(t, T) = F(t; t, T)$.

1.2 Bonds

The zero coupon bond is the most elementary product in the interest rate market. It is a contract which guarantees the holder to be paid out 1[1] at the maturity date T. The price at time t of a zero coupon bond with maturity T is denoted by $p(t, T)$. From the definition it follows that $p(T, T) = 1$. Zero coupon bonds provide no payoff before time T and are therefore very useful instruments for modelling purposes. Zero coupon bonds are also very useful for discounting. Suppose we know we will receive an amount of K at a future time T, then the present value (at time t) of this cash flow is $p(t, T)K$.

The spot rate $R(t, T)$ can be derived from the value of the zero coupon bond. Suppose we invest 1 at time t. If we invest it in a zero coupon bond, we will receive $1/p(t, T)$ at time T. If we invest it at the spot rate, it will pay us $1 + (T - t)\, R(t, T)$. These two results have to be equal, so we get

$$R(t, T) = \frac{1 - p(t, T)}{(T - t)\, p(t, T)}.$$

In the same way we can compute the relation between bond prices and forward rates $F(t, S, T)$, see section (1.4).

In contrast to a zero coupon bond, a coupon bond does pay out a coupon at intermediary points in time. A fixed coupon bond pays out a predetermined coupon whereas a floating rate bond pays out an amount dependent on the market interest rate. Coupon bearing bonds are much more actively traded on the market, but are less useful for the modelling of interest rates. They can be expressed in terms of a portfolio of zero coupon bonds.

1.3 Forward Rate Agreement

A Forward Rate Agreement (FRA) is a contract to let a certain, predetermined, interest rate R, over some future period $[T_{i-1}, T_i]$, act on a prespecified principal K. The lender pays K to the borrower at T_{i-1} and receives $K(1 + \alpha_i R)$ at T_i, where $\alpha_i = T_i - T_{i-1}$. The cash flows for the other party, the borrower, are of course opposite to these. The value of this contract for the lender at time $t < T_{i-1}$ is given by:

$$\mathrm{FRA}(t) = K\left[p(t, T_i)(1 + a_i R) - p(t, T_{i-1})\right]. \tag{1.1}$$

[1] payoffs and values will not be denoted in a certain currency. We will not consider cross-currency products, so everything will be denoted in the same currency.

1.4 Forward Libor rates

In the context of forward rates (section 1.1), usually forward Libor rates are used. LIBOR means London Interbank Offer Rate. It is the forward rate offered by banks to other banks on Eurocurrency deposits. The corresponding bid rate is call LIBID.

First define a Tenor structure, a set of dates:

$$0 \leq T_0 < T_1 < \cdots < T_N.$$

The year fraction between two subsequent dates T_{i-1} and T_i is defined by α_i (usually called the tenor or day count fraction). We will not worry about day count conventions and use

$$\alpha_i = T_i - T_{i-1}.$$

It is assumed that there exist a zero coupon bond $p(t, T_i)$ bond for each maturity T_i. Now it is possible to define the forward Libor rate $L_i(t)$ $(1 \leq i \leq N)$ as the forward rate between two tenor dates:

$$L_i(t) = F\left(t, T_{i-1}, T_i\right).$$

So it is the interest rate that can be contracted at time t for the period $[T_{i-1}, T_i]$, where $t \leq T_{i-1}$, without any costs.

If we compare this with the definition of the Forward Rate Agreement, we can see that $L_i(t)$ is the interest rate R that makes the value of the FRA equal to zero. Solving this from equation (1.1) gives:

$$1 + \alpha_i L_i(t) = \frac{p(t, T_{i-1})}{p(t, T_i)}, \tag{1.2}$$

from which the forward Libor rate can be defined as:

$$L_i(t) = \frac{1}{\alpha_i}\left(\frac{p(t, T_{i-1})}{p(t, T_i)} - 1\right), \tag{1.3}$$

or, alternatively

$$L_i(t) = \frac{p(t, T_{i-1}) - p(t, T_i)}{\alpha_i p(t, T_i)}. \tag{1.4}$$

For $t = T_{i-1}$ the forward Libor rate is equal to the simply-compounded spot rate with maturity T_i:

$$R(T_{i-1}, T_i) \equiv L_i(T_{i-1}) = \frac{1 - p(T_{i-1}, T_i)}{\alpha_i p(T_{i-1}, T_i)}. \tag{1.5}$$

Therefore $L_i(T_{i-1})$ is also called the Libor rate, in contrast to the forward Libor rate $L_i(t)$ $(t < T_{i-1})$. In this thesis I will use the following short-hand notation

$$L_i \equiv L_i(T_{i-1}).$$

For the ease of notation, I will sometimes refer to $L_i(t)$ as the Libor rate, instead of the forward Libor rate, when it is clear from the context that we mean the forward rate.

As we will see in the following sections, other products are defined in terms of (forward) Libor rates. Therefore it is no surprise that it would be useful if we could make an interest rate model that describes the dynamics of the Libor rates. This is exactly the aim of the Libor Market Model.

1.5 Swaps

An interest rate swap is a contract to exchange a set of floating rate payments (floating leg) for a set of fixed payments (fixed leg). The floating leg usually is a payment of the Libor rate over a specified amount. The fixed leg is a fixed rate (also called the strike) over the same amount. There are several versions of interest rate swaps, but here the *forward swap settled in arrears* will be used. The owner of a receiver swap will receive the fixed rate and pay the floating rate. For a payer swap floating is received and fixed is paid.

Denote the principal by K and the strike by R_S. A $T_n \times (T_N - T_n)$ swap is a swap with maturity T_n and tenor $T_N - T_n$. At each reset date T_i, $n \leq i \leq N-1$ the Libor rate $L_{i+1}(T_i)$ is observed in the market. For a payer swap, at T_{i+1} a payment of $K\alpha_{i+1}R_S$ has to be made and an amount of $K\alpha_{i+1}L_{i+1}$ is received. See figure 1.2 for an example of the cash flows of a swap.The net cash flow at T_{i+1} to the holder is thus

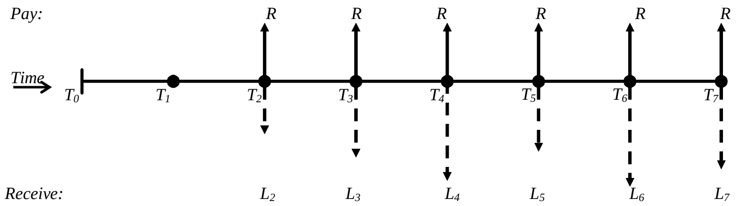


Figure 1.2: Cash flows of a $T_1 \times (T_7 - T_1)$ payer swap $(K = 1, \alpha_i = 1)$

$$K\alpha_{i+1}\left(L_{i+1} - R_S\right).$$

We can discount the net cash flow, to get the value at time t of this cash flow (using equation 1.3):

$$Kp(t, T_i) - K\left(1 + \alpha_{i+1}R_S\right)p(t, T_{i+1}).$$

Note that this is equal to -1 times value of an FRA (equation 1.1). Summing the value of all the payments at $T_{n+1}, \ldots, T_N$ we find the value for the $T_n \times (T_N - T_n)$ payer swap:

$$\begin{aligned} \mathrm{PS}_n^N(t) &= K\sum_{i=n}^{N-1}\left[p(t,T_i) - \left(1+\alpha_{i+1}R_S\right)p(t,T_{i+1})\right] \qquad (1.6)\\ &= K\left(p(t,T_n) - p(t,T_N) - RA_n^N(t)\right), \end{aligned}$$

where $A_n^N(t)$ is the annuity, or present value of a basis point (PVBP, because it corresponds to the increase in value of the fixed side of the swap if the swap rate R increases, see equation 1.6):

$$A_n^N(t) = \sum_{i=n+1}^{N} \alpha_i p(t, T_i). \qquad (1.7)$$

The swap rate or forward swap rate $R_n^N(t)$ is defined as the strike R_S for which the value of the swap contract is equal to zero. This can be derived by setting the value of the swap equal to zero:

$$R_n^N(t) = \frac{p(t, T_n) - p(t, T_N)}{A_n^N(t)}. \qquad (1.8)$$

Using this, the value of the payer swap can also be expressed as:

$$\mathrm{PS}_n^N(t) = K\left(R_n^N(t) - R_S\right)A_n^N(t). \tag{1.9}$$

A swap is called at-the-money (ATM) if $R_n^N(t) = R$, so its value is equal to zero. If the value of the swap is positive or negative it is called in-the-money (ITM) or out-of-the-money (OTM) respectively. This is called the moneyness of the swap.

1.6 Caps and floors

A cap is one of the most important plain vanilla options in the interest rate market. A cap is designed to provide insurance against the rate of interest on the floating-rate note rising above a certain level, known as the cap rate. Denote the cap rate by R and the principal by K. We use the same term structure as before. If $L_i(T_{i-1}) > R$, the cap pays the difference between L_i and R and nothing if $L_i < R$. Its payoff at time T_i, $1 \leq i \leq N$ can be written as:

$$K\alpha_i\left[L_i - R\right]^+,$$

where we use the notation $[x]^+ \equiv \max(x, 0)$. The payoff is equal to the payoff of a call option on the Libor rate. So a cap is a set of options, one for each reset date. The N call options are called caplets. The value of the cap is equal to the sum of the values of the caplets:

$$\mathrm{Cap}(t) = \sum_{i=1}^{N} \mathrm{capl}_i(t).$$

Figure 1.3 gives an example of the cash flows from a cap. In the market, the value of a

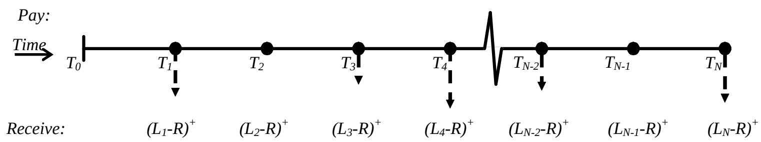


Figure 1.3: Cash flows of a cap ($K = 1$, $\alpha_i = 1$)

caplet is usually determined using Black's (1976) model (see Appendix A.1). This requires the assumption that

$$\ln\left(L_i\left(T_{i-1}\right)\right) \sim \mathcal{N}\left(\ln\left(L_i(t)\right), \overline{\sigma}_i\sqrt{T_{i-1} - t}\right).$$

So $L_i\left(T_{i-1}\right)$ is lognormally distributed. $\overline{\sigma}_i$ is the volatility of $L_i(t)$. Under this assumption, the value of the caplet is given by Black's formula:

$$\mathrm{capl}_i(t) = K\alpha_i p(t, T_i)\left[L_i(t)N\left(d_1\right) - RN\left(d_2\right)\right], \tag{1.10}$$

where

$$\begin{aligned} d_1 &= \frac{\ln\left(L_i(t)/R\right) + \overline{\sigma}_i^2\left(T_{i-1} - t\right)/2}{\overline{\sigma}_i\sqrt{T_{i-1} - t}} \\ d_2 &= d_1 - \overline{\sigma}_i\sqrt{T_{i-1} - t}. \end{aligned}$$

Here $N(\cdot)$ is the standard normal distribution function. Even though this method is widely used to price caplets, the traditional interest rate models do not imply lognormal Libors. Nevertheless, caps (and swaptions) are typically quoted in terms of their Black-implied volatility.

A floor is the opposite of a cap and provides insurance against the rate falling below a certain level. Just like a cap is a collection of caplets, a floor is a collection of floorlets. Using the same notation as above, the net cash flow of a floorlet at T_i is

$$K\alpha_i \left[R - L_i\right]^+ .$$

Its value is given by

$$\text{floorl}_{\,i}(t) = K\alpha_i p(t, T_i) \left[RN(-d_2) - L_i(t)N(-d_1)\right].$$

The moneyness of caplets at time t is defined by the underlying $L_i(t) - R$. When this is equal to zero, larger or smaller than zero, the caplet is at-, in- or out-of-the-money respectively. For floorlets the same holds for the underlying $R - L_i(t)$.

Digitals are very similar to caps and floors. Instead of paying $L_i - R$, a digital cap pays 1 when $L_i > R$ and nothing otherwise. The value of a digital caplet is given by:

$$\text{digicapl}_{\,i}(t) = K\alpha_i p(t, T_i) N(d_2). \tag{1.11}$$

A digital floor pays the opposite of the digital cap: $1_{\{L_i < R\}}$. Its value is given by:

$$\text{digifloorl}_{\,i}(t) = K\alpha_i p(t, T_i) N(-d_1),$$

where d_1 and d_2 are the same as for the cap.

1.7 European swaptions

A swaption is an option on a plain vanilla swap (see section 1.5). It gives the holder the right (but not the obligation) to enter into a certain interest rate swap, called the underlying, at a certain time T_n, the expiry date of the swaption. A receiver swaption gives the right to enter into a receiver swap, a payer swaption to enter a payer swap.

At expiry, if the value of the underlying is positive, the swaption will be exercised, so the holder of the option will receive the swap whose value is given by equation (1.9). If the value of the underlying is negative, the swaption will not be exercised, so the holder will receive nothing. Therefore the value at expiry T_n of the $T_n \times (T_N - T_n)$ payer swaption is given by:

$$\left[\text{PSN}_{\,n}^{\,N}(T_n)\right]^+ = K\left[R_n^N(T_n) - R\right]^+ A_n^N(T_n).$$

The value of the European swaption can be obtained by using Black's formula (see also appendix A.1). If we take $A_n^N(T_n)$ as the numeraire, the above formulation shows that the payer swaption is just a call option on R_n^N with strike R. If we assume $R_n^N(T_n)$ is lognormally distributed with

$$\ln\left(R_n^N(T_n)\right) \sim \mathcal{N}\left(\ln\left(R_n^N(t)\right), \sigma_{n,N}\sqrt{T_n - t}\right),$$

then Black's formula gives the following value of the payer swaption:

$$\text{PSN}_n^N(t) = KA_n^N(t)\left[R_n^N(t)N(d_1) - RN(d_2)\right], \tag{1.12}$$

where

$$
\begin{aligned}
d_1 &= \frac{\ln\left(R_n^N(t)/R\right) + \sigma_{n,N}^2 \left(T_n - t\right)/2}{\sigma_{n,N}\sqrt{T_n - t}} \\
d_2 &= d_1 - \sigma_{n,N}\sqrt{T_n - t}.
\end{aligned}
$$

This is also the method which is used in the market to price swaptions. However, it can be shown (see section 3.4) that when $L_i(T_{i-1})$ is lognormally distributed, $R_n^N(T_n)$ is not, and vice versa. That means that the pricing of caplets and floorlets is not consistent with the pricing of European swaptions. We will come back to this issue in section 3.4.

2 Introduction to interest rate models

With the growth of the interest rate derivatives market, it became important to develop models to price these products. Since the first models in the 1970's, new, more realistic models were developed to incorporate more information of the interest rate market and to be able to price more exotic derivatives.

After the breakthrough in stock option pricing theory by Black and Scholes (1973) and Merton (1973), the valuation of interest rate derivatives started with Black's (1976) model. This was used to value caps, options on bonds and European swaptions. This model assumes that the probability distribution of an interest rate, bond price or another variable at a future time is lognormal. This model is still widely used for some products, but has important limitations. It is not consistent; if a bond price is lognormally distributed, the interest rate is not, so we cannot use this model to price bond options as well as caps. Furthermore, it only gives the distribution of a single underlying at a single moment and does not give any information about the development of interest rates through time or the correlations between underlyings. Therefore they can not be used to value other products, whose value depends on more than a single date.

This led to the development of term structure models. These models give a description of the risk-neutral evolution of interest rates through time. The first term structure models were short rate models. These models describe the development of the short rate $r(t)$ (see also section 1.1). From this, it is possible to define the spot rate, the interest rate over a period of time $[t, T]$, by

$$R(t,T) = \frac{1}{T-t}\int_t^T r(s)ds.$$

The most important models of this type are the ones developed by Vasicek (1977) and Cox, Ingersoll and Ross (1985). For example, in the Cox, Ingersoll and Ross (CIR) model, the process for $r(t)$ is:

$$dr = a\,(b-r)\,dt + \sigma\sqrt{r}dz,$$

where a, b are constant, σ is the volatility of the short rate and dz is a Brownian motion. Later, these models were extended to make them consistent with the initial term structure (by Ho and Lee, 1986; Hull and White, 1990; Black, Derman and Toy, 1990).

All these models provide ways to price derivatives when Black's model is inappropriate and are easy to implement. However, they still have some important limitations. They only describe one rate (the short rate) and therefore the interest rates are driven by only one source of uncertainty (dz). That implies all interest rates will be affected by a single factor. In reality however, short term interest rates are usually affected by different events than long term rates. It was tried to solve these problems by adding an extra factor, leading to two factor models (Duffie and Kan, 1996; Hull and White, 1994).

Meanwhile, Heath, Jarrow and Morton (1992) had developed a total different method to model interest rates. Instead of focussing on the short rate, they modelled the instantaneous forward rate $f(t,T)$ (see section 1.1). The HJM model assumes that for every fixed $T > 0$ the instantaneous forward rate have the following dynamics:

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)dW,$$

where $\alpha(t,T)$ and $\sigma(t,T)$ are adapted processes and dW is a D-dimensional Wiener-process. The model provides a process for every instantaneous forward rate, giving much more flexibility than the one- or two-factor models. The drawback of the HJM model is that it is expressed in terms of instantaneous forward rates, which are not observable in the market. That makes the model harder to calibrate. By the way, the same problem applies to the short rate models: there is no such thing as a short rate in the market.

This leaded to the development of a similar model, now defined in terms of the forward rate $F(t, T_1, T_2)$, instead of the instantaneous forward rate. These forward rates are traded in the market. Furthermore, the model is consistent with Black's formula for the pricing of caps (equation 1.10), which is still the usual way these product are priced.

Because Libor rates are the most actively used forward rates, the model was named the Libor Market Model (LMM). The model has been introduced by Brace, Gatarek and Musiela (1997), after who it is sometimes called the BGM model, Jamshidian (1997) and Miltersen, Sandmann and Sondermann (1997). Since its introduction it has become a very important model for the pricing of a wide range of interest rate products, most notably the so-called Libor Exotics, whose payoff depends on Libor rates.

3 Libor Market Model

In this chapter the Libor Market model (LMM) will be described. Like other interest rate models, the objective of the LMM is to provide a model of the dynamics of the evolution of interest rates to price non-standard interest rate derivatives in such a way that it is consistent with the market prices of other basic (plain-vanilla) products. The most important plain-vanilla products have been defined in Chapter 1. The LMM is exactly consistent with the use of Black's formula for the pricing of caplets by assuming Libor rates are lognormally distributed. The model specifies the (continuous) dynamics of the forward Libor rates for discrete maturities, being the tenor dates. Because the model is consistent with the valuation of caplets, it turns out to be easy to calibrate it to market data (see also Chapter 4).

The next section defines the LMM. Section 3.2 gives the drift of the forward Libors under different measures and in Section 3.3 an alternative way to formulate the LMM is given. The valuation methods for pricing caps and swaptions as described in Chapter 1, are inconsistent. When Libor rates are lognormally distributed, swap rates are not. Section 3.4 shows this and derives the volatility of the swap rates under the LMM.

This chapter frequently uses results from basic financial calculus, like risk neutral valuation, equivalent martingale measures, Girsanov's theorem, Radon-Nikodym derivatives. We will not explain these results in this thesis, but for more information the reader can find useful references in appendix A.

3.1 Definition

We will use the definition of the forward Libor rates from Section 1.4. So we take a tenor structure

$$0 \leq T_0 < T_1 < \cdots < T_N,$$

with the tenors $\alpha_i = T_i - T_{i-1}$ and Libor rates $L_i(t) = F(t, T_{i-1}, T_i)$. The Libor Market Model (LMM) assumes that the forward Libor rates $L_i(t)$ are instantaneously lognormally distributed. This means L_i has the following dynamics:

$$dL_i(t) = \ldots dt + L_i(t)\zeta_i'(t)d\mathbf{W}(t),$$

where $\mathbf{W}(t)$ is a D-vector of independent standard Wiener processes on $(\Omega, \mathcal{F}_t, \mathbb{P})$. D is the number of factors ($D \leq N$). If $D = N$, there is one source of uncertainty per Libor rate. $\zeta_i(t)$ is a D-vector containing the volatility of L_i. The d'th component of the vector is the volatility of $L_i(t)$ corresponding to the d-th factor. We will assume $\zeta_i(t)$ is a deterministic function of time t. When $\zeta_i(t)$ is not deterministic but follows a stochastic process, we have a so-called stochastic volatility model.

As will be shown in Section 3.2, the drift depends on the L_i's and we find the dynamics to have the following form:

$$\begin{aligned} dL_i(t) = \mu_i\left(\mathbf{L}(t), t\right) L_i(t)dt + \zeta_i'(t)L_i(t)dW(t) \\ t \leq T_{i-1}, 1 \leq i \leq N \end{aligned} \tag{3.1}$$

Here μ_i is the drift of the i'th Libor rate. $\mathbf{L(t)}$ is the vector of forward Libors $L_i(t)$. Note that this model is only defined for $t \leq T_{i-1}$, because L_i resets at time T_{i-1}, the volatility and drift are equal to 0 for $t \geq T_{i-1}$.

Because μ_i depends on $\mathbf{L}$, the Libors are no longer lognormally distributed. Note that $L_i(t)$ is of course still instantaneous lognormal, due to the formulation (3.1). However, for Black's formula $L_i(T_{i-1})$ conditioned on $\mathcal{F}_t$ should be lognormal, which is not the case. Fortunately, it is possible to find a different measure, under which L_i *is* lognormally distributed.

First denote by $\mathbb{Q}^i$ the Martingale measure with $p(t, T_i)$ as numeraire. Denote by $\mathbf{W}^i$ a D-dimensional $\mathbb{Q}^i$-Wiener process. We will use the results of Harrison and Kreps (1979) that, in a market where there is no arbitrage, for any given strictly positive numeraire security whose price is $g(t)$, there exists a measure for which $f(t)/g(t)$ is a martingale for all security prices $f(t)$. From equation (1.2) it follows that $1 + \alpha_i L_i(t)$ is a martingale under the measure $\mathbb{Q}^i$. So also the forward Libor rate is a martingale under this measure (also known as the natural measure):

$$dL_i(t) = \zeta_i'(t) L_i(t) d\mathbf{W}^i(t), \qquad 1 \leq i \leq N. \tag{3.2}$$

The solution is given by:

$$L_i(T) = L_i(t) e^{Y_i(t,T)}, \tag{3.3}$$

where $Y_i(t, T)$ is normally distributed with mean m_i and variance $\overline{\sigma}_i^2 (T - t)$ given by:

$$\begin{aligned} m_i(t, T) &= -\frac{1}{2}\overline{\sigma}_i^2, \\ \overline{\sigma}_i^2 &= \frac{1}{T-t}\int_t^T |\zeta_i(s)|^2 ds. \end{aligned} \tag{3.4}$$

Under its natural measure, L_i just follows geometric Brownian motion and thus is lognormally distributed, which is exactly what we wanted to obtain to be able to use Black's model for caplets. Therefore caplets can be priced exactly by Black's formula (1.10). For this reason $\overline{\sigma}_i$ (the term volatility) is also called the Black (caplet) volatility.

3.2 Drift under different measures

Note that in equation (3.2) each forward Libor rate is a martingale under its own natural measure, but not under the same measure! Now we will derive the dynamics of the forward Libor rates under a single measure, being the terminal measure $\mathbb{Q}^N$. This is the measure with $p(t, T_N)$ as numeraire. To apply this change of measure, we need to find the Girsanov kernel (see e.g. Björk, 2004). The measures $\mathbb{Q}^i$ and $\mathbb{Q}^{i-1}$ are absolutely continuous with respect to each other, and the Radon-Nikodym derivative η_i^{i-1} is given by:

$$\eta_i^{i-1}(t) = \frac{d\mathbb{Q}^{i-1}}{d\mathbb{Q}^i} = \frac{p(T_0, T_i)}{p(T_0, T_{i-1})} \cdot \frac{p(t, T_{i-1})}{p(t, T_i)} = A_i (1 + \alpha_i L_i(t)),$$

where $A_i = \frac{p(T_0, T_i)}{p(T_0, T_{i-1})}$. From equation (3.2) follows :

$$\begin{aligned} d\eta_i^{i-1}(t) &= A_i \alpha_i \zeta_i'(t) L_i(t) d\mathbf{W}^i(t) \\ &= A_i (1 + \alpha_i L_i(t)) \frac{\alpha_i L_i(t)}{(1 + \alpha_i L_i(t))} \zeta_i'(t) d\mathbf{W}^i(t) \\ &= \eta_i^{i-1}(t) \frac{\alpha_i L_i(t)}{(1 + \alpha_i L_i(t))} \zeta_i'(t) d\mathbf{W}^i(t). \end{aligned}$$

This shows the Girsanov's kernel is given by:

$$\frac{\alpha_i L_i(t)}{(1 + \alpha_i L_i(t))} \zeta_i(t).$$

Girsanov's theorem now shows how to change measures:

$$d\mathbf{W}^i(t) = \frac{\alpha_i L_i(t)}{(1+\alpha_i L_i(t))}\zeta_i(t)dt + d\mathbf{W}^{i-1}(t).$$

We can repeat this to find for the terminal measure:

$$d\mathbf{W}^N(t) = \sum_{k=i+1}^{N} \frac{\alpha_k L_k(t)}{(1+\alpha_k L_k(t))}\zeta_k(t)dt + d\mathbf{W}^{i}(t).$$

This shows the $\mathbb{Q}^N$ dynamics of the forward Libor rate:

$$dL_i(t) = -L_i(t)\left(\zeta_i'(t)\sum_{k=i+1}^{N} \frac{\alpha_k L_k(t)}{(1+\alpha_k L_k(t))}\zeta_k(t)\right)dt + \zeta_i'(t)L_i(t)d\mathbf{W}^N(t). \tag{3.5}$$

In the same way, we can compute the drift for every numeraire bond. Let μ_i^j be the drift of the ith Libor rate under the martingale measure $\mathbb{Q}^j$ (i.e. with $p(t, T_j)$ as numeraire). Then, for $t \leq \min(T_i, T_{j-1})$:

$$\mu_i^j(\mathbf{L}(t), t) = \begin{cases} \zeta_i(t) \displaystyle\sum_{k=j+1}^{i} \tfrac{\alpha_k L_k(t)}{(1+\alpha_k L_k(t))}\zeta_k(t) & \text{if } i > j \\ 0 & \text{if } i = j \\ -\zeta_i(t) \displaystyle\sum_{k=i+1}^{j} \tfrac{\alpha_k L_k(t)}{(1+\alpha_k L_k(t))}\zeta_k(t) & \text{if } i < j. \end{cases} \tag{3.6}$$

Finally we will derive the risk neutral dynamics. These are obtained by using the risk-neutral bank account $B(t)$ as the numeraire. Usually B is defined through

$$\begin{aligned} dB(t) &= r(t)B(t)dt, \\ B(T_0) &= 1, \end{aligned}$$

where $r(t)$ is the short rate. In the LMM the short rate is not defined and therefore it is more natural to take as bank account a portfolio consisting of a bond with the shortest maturity. At the maturity of the bond the money is reinvested in the following bond with shortest time to maturity. So at time T_0 we buy a zero-coupon bond with maturity T_1. At time T_1 this is worth $\frac{1}{p(T_0,T_1)}$, which we reinvest in a bond with maturity T_2, etcetera. Define $m(t) = \min(i : T_i \geq t)$ as the next reset moment, so $T_{m(t)-1} < t \leq T_{m(t)}$. The value of this portfolio at time t is given by.

$$B(t) = \frac{p(t, T_{m(t)})}{\prod_{j=1}^{m(t)} p(T_{j-1}, T_j)}, \tag{3.7}$$

where we use $\prod_{j=1}^{0} .. = 1$. The corresponding martingale measure $\mathbb{Q}^B$ is called the risk-neutral measure or the spot Libor measure. Using this definition, the Radon-Nikodym derivative for the change of measure from $\mathbb{Q}^i$ to $\mathbb{Q}^B$, is given by:

$$\eta_i^B = \frac{p(T_0, T_i)}{B(T_0)}\frac{B(t)}{p(t, T_i)} = \frac{p(T_0, T_i)p(t, T_{m(t)})}{p(t, T_i)\prod_{j=1}^{m(t)} p(T_{j-1}, T_j)}.$$

If we now take $i = m(t)$, we get:

$$\eta^B_{m(t)} = \frac{p(T_0, T_{m(t)})p(t, T_{m(t)})}{p(t, T_{m(t)}) \prod_{j=1}^{m(t)} p(T_{j-1}, T_j)} = \frac{p(T_0, T_{m(t)})}{\prod_{j=1}^{m(t)} p(T_{j-1}, T_j)}.$$

This is just a constant (because it resets at $T_{m(t)-1}$, which is smaller than t. That means the Girsanov Kernel is zero, and therefore the dynamics are the same as under the $\mathbb{Q}^i$ dynamics with $i = m(t)$. The drift follows from equation (3.6) (just take $j = m(t)$ and note that $i \geq m(t)$):

$$\mu^B_i\left(\mathbf{L}(t), t\right) = \zeta_i'(t) \sum_{k=m(t)+1}^{i} \frac{\alpha_k L_k(t)}{(1 + \alpha_k L_k(t))} \zeta_k(t).$$

3.3 Alternative Formulation

It is also possible to use, in contrast to the formulation in Section 3.1, an alternative formulation with scalar Wiener processes, i.e. with one Wiener process for each Libor rate. Then the Wiener processes are no longer independent. In the formulation above, there were several independent Wiener processes, each influencing all Libor's. These factors are risk-factor-specific and can for example be shifts in the yield curve, changes in its slope or curvature, etc. In the formulation below, the Wiener processes are forward-rate-specific. First define the volatility of the Libor rate by:

$$\sigma_i(t) = |\zeta_i(t)| = \sqrt{\zeta_i' \zeta_i}. \tag{3.8}$$

Now define the following scalar Wiener process:

$$dW_i = \frac{1}{\sigma_i(t)} \zeta_i' d\mathbf{W}.$$

These are correlated scalar Wiener processes. The correlation ρ_{ij} is defined by by $dW_i(t)dW_j(t) = \rho_{ij}dt$. Because the original $\mathbf{W}$ is a vector of independent Wiener processes, there holds $d\mathbf{W}d\mathbf{W}' = Idt$ (I is the identity matrix). So the the correlation can be computed as:

$$\begin{aligned} \rho_{ij}dt &= dW_i(t)dW_j(t) = \frac{1}{\sigma_i(t)}\zeta_i' d\mathbf{W} \frac{1}{\sigma_j(t)} d\mathbf{W}'\zeta_i \\ &= \frac{1}{\sigma_i(t)\sigma_j(t)} \zeta_i'\zeta_j dt = \frac{\zeta_i'\zeta_j}{\sigma_i(t)\sigma_j(t)} dt. \end{aligned}$$

From this, it also follows that

$$\zeta_i'\zeta_j = \sigma_i(t)\sigma_j(t)\rho_{ij}(t). \tag{3.9}$$

Substituting everything into equation (3.1) yields:

$$\begin{aligned} dL_i(t) &= \mu_i\left(\mathbf{L}(t), t\right) L_i(t)dt + \sigma_i(t) L_i(t) dW_i(t) \\ dW_i(t)dW_j(t) &= \rho_{ij}(t)dt. \end{aligned} \tag{3.10}$$

This formulation is more intuitive than the original, because there is just one scalar volatility function for each Libor rate. However it may be less clear how a lower number of driving factors is implemented. For $D < N$ factors the correlation matrix will have rank D.

For example, under the terminal measure, μ_i can be obtained by substituting equation (3.9) into equation (3.5):

$$\mu_i\left(\mathbf{L}(t), t\right) = -\sigma_i(t) \sum_{k=i+1}^{N} \frac{\alpha_k L_k(t)}{(1+\alpha_k L_k(t))} \sigma_k(t)\rho_{ik}. \tag{3.11}$$

For the rest of this thesis, we will use the notation from this section.

3.4 Distribution of the swap rate

As noted in section 1.7, Black's formula is usually used to value (and quote prices of) swaptions. This requires the assumption that the swap rate is lognormally distributed (under the appropriate numeraire). As will be shown below, in the LMM the swap rate is not lognormal. Fortunately, it turns out that this inconsistency does not lead to big problems, because the value can be approximated very accurately (Jäckel and Rebonato, 2003).

It is also possible to make another model, in contrast to the LMM, which assumes the swap rates to be lognormally distributed: the swap market model. Of course, then it is no longer possible to value caplets analytically. Moreover, the swap market model leads to a more complicated drift functions and is harder to calibrate and therefore the LMM is preferred.

We will now derive the swaption volatility approximation from Jäckel and Rebonato (2003). Another way to find the same results can be found in (Hull and White, 1999), although less straightforward. It is easy to see that the swap rate (see equation 1.8) can be rewritten as:

$$R_n^N(t) = \sum_{i=n+1}^{N} w_i(t) L_i(t), \tag{3.12}$$

where the weights w_i are defined by:

$$w_i(t) = \frac{\alpha_i p(t, T_i)}{A_n^N(t)}.$$

This can easily be checked by substituting w_i and (1.4) into equation (3.12). We see that, if L_i is lognormally distributed, R_n^N is not. By applying Itô's lemma to $R_n^N(t)$, we can find the volatility of the swap rate:

$$\begin{aligned} \sigma^2_{n,N}(t) &= \frac{\sum_j \sum_k \left[\partial R_n^N / \partial L_j\right] \left[\partial R_n^N / \partial L_k\right] L_j(t) L_k(t) \rho_{jk}(t) \sigma_j(t) \sigma_k(t)}{\left[\sum_i w_i L_i(t)\right]^2} \\ &= \sum_j \sum_k \zeta_{jk}(t) \rho_{jk}(t) \sigma_j(t) \sigma_k(t), \end{aligned} \tag{3.13}$$

where

$$\zeta_{jk}(t) = \frac{\left[\partial R_n^N / \partial L_j\right] \left[\partial R_n^N / \partial L_k\right] L_j(t) L_k(t)}{\left[\sum_i w_i L_i(t)\right]^2}.$$

If we look at equation (3.12) it is tempting to compute the derivatives as:

$$\frac{\partial R_n^N}{\partial L_i} = w_i. \tag{3.14}$$

However, this is not correct, because w_i depends on L_i and therefore is stochastic. Nevertheless, to approximate the volatility, we can assume that equation (3.14) is correct. A more precise estimation is given below.

To be able to compute the Black-volatility, that can be used in equation (1.12), we need to compute:

$$\begin{aligned}\overline{\sigma}_{n,N}^2 (T_n - t) &= \int_t^{T_n} \sigma_{n,N}^2(u)du \\ &= \int_t^{T_n} \zeta_{jk}(u)\rho_{jk}(u)\sigma_j(u)\sigma_k(u)du.\end{aligned}$$

Brace and Womersley (2000) show that it is possible to approximate $\zeta_{jk}(u)$ $(u \geq t)$ accurately by $\zeta_{jk}(t)$, because $\zeta_{jk}(u)$ is a Martingale and has a relatively low variance (compared to the variance of the Libor rates). That implies:

$$\overline{\sigma}_{n,N}^2 (T_n - t) = \zeta_{jk}(t) \int_t^{T_n} \rho_{jk}(u)\sigma_j(u)\sigma_k(u)du. \tag{3.15}$$

When the volatilities and correlations are deterministic, the integral can be computed. This gives an approximation of the swap rate volatility.

In the approximation above we used the wrong assumption from equation (3.14). It is also possible to use the correct derivative. This leads to the following value for the coefficients (suppressing the dependency on t of L_i, A_i, B_i):

$$\begin{aligned}\zeta_{ij}(t) &= \left[\frac{p(t,T_{i+1})\alpha_{i+1}L_{i+1}}{A_n} + \frac{(A_nB_i - A_iB_n)\,\alpha_{i+1}L_{i+1}}{A_nB_n\,(1+\alpha_{i+1}L_{i+1})}\right] \\ &\quad \cdot \left[\frac{p(t,T_{i+1})\alpha_{i+1}L_{i+1}}{A_n} + \frac{(A_nB_j - A_jB_n)\,\alpha_{i+1}L_{i+1}}{A_nB_n\,(1+\alpha_{i+1}L_{i+1})}\right],\end{aligned}$$

where

$$\begin{aligned}A_i &= \sum_{j=i+1}^{N} p(t,T_j)\alpha_jL_j, \qquad n \leq i \leq N-1, \\ B_i &= \sum_{j=i+1}^{N} p(t,T_j)\alpha_j, \qquad n \leq i \leq N-1.\end{aligned}$$

Jäckel and Rebonato (2003) show this leads to a clear improvement over the original approximation. In this thesis, when speaking about approximating the volatility of the swap rate, we are referring to Jäckel and Rebonato's (2003) approximation.

4 Calibration

Before we can use the Libor Market model to price derivatives, we have to determine the parameters of the model, i.e. calibrate it. This chapter describes the most important issues in the calibration of the LMM. Calibration of the Libor Market model contains a whole area of research on its own, so it is impossible to discuss every topic. For more information on calibration, see for example (Rebonato, 1999; Rebonato, 2002; Brigo and Mercurio, 2001).

4.1 Objective

In the interest rate market, bonds, swaps, caps and swaptions are traded frequently. Therefore the prices of these products are very accurate. The Libor Market Model does not give prices for those products, but uses them as input, such that the model gives prices for other products that are consistent with the prices of plain-vanilla products. That means that we have to calibrate the model to these prices, such that if we value a cap or a swaption with the model, we retrieve the prices quoted in the market. If we do this, we ensure that the prices of exotic derivatives will be consistent with the market prices of bonds, caps, and swaptions.

In the LMM, bonds and caps are priced exactly (analytically) and therefore we also want the calibration in such a way that the model is exactly fitted to these prices. Swaptions however, can not be valued exactly analytically, even though very good approximations exist (see also Section 3.4). Therefore we also do not require swaption prices to be exactly consistent with the LMM, but we do want a good fit.

Under the alternative formulation (equation 3.10), the evolution of the Libor rate is fully determined by three elements:

- the initial Libor rates: $L_i(T_0)$,
- the volatility functions of the Libor rates: $\sigma_i(t)$, $(T_0 \leq t \leq T_{i-1})$,
- the correlations between the different factors: $\rho_{ij}(t)$, $(T_0 \leq t \leq T_{i-1})$,

In the following sections, these will be addressed separately. Once these variables have been set, the whole model is determined and can be used to price derivatives.

4.2 Initial Libors

The first task is easy. At time T_0 the Libor rates can simply be observed in the market, which is the reason why these models are called market models. We just put $L_i(T_0)$ in the model equal to these market Libors. Because the Libors are defined in terms of zero coupon bonds, this automatically ensures consistency with bond prices. Usually Libor rate quotes and bond prices do not exist for every tenor date T_i. In these cases we can interpolate between existing quotes, to estimate the forward rate.

4.3 Volatility

For the calibration of volatility we have to determine the function $\sigma_i(t)$ for $2 \leq i \leq N$ and $T_0 \leq t \leq T_{i-1}$. This task is much harder than the calibration to the Libor rates, because we do not only have to estimate the current volatility $\sigma_i(T_0)$ but also the future volatility.

The volatility $\sigma_i(t)$ as a function of t is called the volatility structure. Estimating future volatility is as hard as estimating future interest rates and usually the best we can do is take the current volatilities as estimate for the future, as will be discussed below.

Because we want Black's formula to be exactly consistent with the LMM, we have to choose the Libor volatilities consistent with the Black volatilities (see also equations 3.4 together with 3.8 and 1.10) :

$$\overline{\sigma}_i^2 = \frac{1}{T_{i-1} - T_0} \int_{T_0}^{T_{i-1}} \sigma_i^2(u) du. \tag{4.1}$$

The 'true' values of $\overline{\sigma}_i^2$ can be computed from the values of caplets in the market. If we choose the volatility function $\sigma_i(t)$ such that equation (4.1) holds, the model value of caplets will be the same as the market value.

Besides caps (and floors), we also want the model to be consistent with European swaption prices. As we have seen in section 3.4, the LMM is not exactly consistent with Black's formula for swaption, but there exists a very good approximation. Because swaptions are traded very actively, we also want the LMM to be consistent with these prices. Therefore we have to make sure that equation (3.15) holds approximately.

Because the system of equations is still highly underdetermined, it is possible to impose some structure on the volatilities. This is desirable, because it prevents the volatility functions from being very irregular. A useful property we would like the volatility structure to have is that it ensures the volatility structure as a function of the time to maturity is (almost) constant through time, i.e. time-homogeneous. The rationale behind this is that we do not know anything about the future development of the volatility term structure ($\sigma_i(t)$ for $t > T_0$) and therefore we want the current structure as expected future volatility term structure. This can be incorporated in the model by using a structure of the form $\sigma_i(t) \sim \sigma(T_{i-1} - t)$. The two most important methods to incorporate this are explained in the following two subsections.

4.3.1 Piecewise constant volatilities

Assume the volatility function is constant between two reset dates. As long as we are only modelling the forward Libor rates at the tenor dates this does not lead to any loss in generality (as we will see in Chapter 7).

$$\sigma_i(t) = \widetilde{\sigma}_{ik}, \qquad (T_{k-1} \le t \le T_k, \quad 1 \le k < i) \,.$$

At $k = i$ the Libor rate matures, so for $k \ge i$ the Libor L_i is no longer stochastic and the volatility is not defined (or equal to 0). To ensure the volatility structure does not change too much through time, we can assume the following structure:

$$\sigma_i(t) = \Phi_i \psi_{i-k}.$$

Here $i - k$ is the number of time periods until maturity, with Φ_i and ψ_j both being constants.

Lets first assume that $\Phi_i = 1$ for all i. Then the volatility structure is constant through time. The cap with maturity T_k can be used to determine ψ_{i-k}, ensuring a perfect fit with caplet prices. However, it is not possible to incorporate swaption prices.

If we do not restrict ourselves to $\Phi_i = 1$, we can obtain a better fit using a two step procedure: first try to obtain a good fit of swaptions and caplets by determining ψ_{i-k}, for example by using some kind of least squares optimization. Now, caplet prices are no longer exactly fitted. This can however be obtained by choosing Φ_i correctly. Because the ψ's already ensured a close fit, Φ_i will not deviate much from 1, so the volatility is still almost time-independent. There is one Φ_i for every caplet, so an exact fit is always easily obtained.

4.3.2 Parametrization

Instead of choosing for piecewise-constant instantaneous volatilities, it is also possible to choose a parametric form for the volatility structure. In the market, the graph of the forward volatility typically is humped-shaped, like in figure 4.1. Rebonato (1999) gives

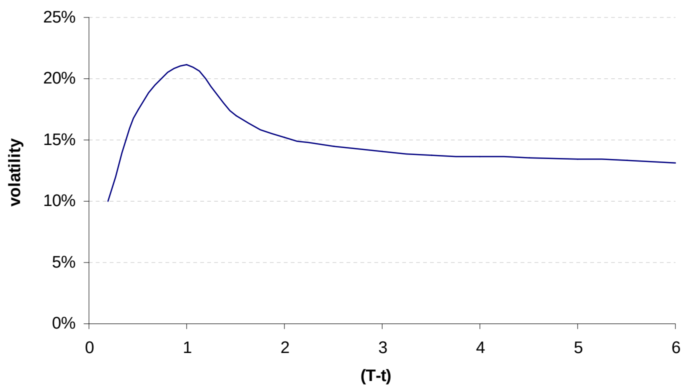


Figure 4.1: typical forward volatility structure

the following possible explanation for this phenomenon. The volatility is caused by uncertainty about future interest rate. For short maturities, up to 6 months or a year, the interest rate is mainly influenced by the monetary authorities (central banks), who communicate their plans well in advance, leading to a low volatility. The longer maturity rates are mainly influenced by the long-term inflation expectations. The monetary authorities usually have a clear inflation target. Therefore their inflation target is usually a good estimate of future inflation, leading to less uncertainty in long-term interest rates. Because there is less consensus about the interest rates for maturities between 6 and 18 months, these have higher volatilities. Note that sometimes, when there is no consensus about short rate actions of the central banks, these volatilities are higher than usual, which is also confirmed by market observations (no longer a hump, but volatility is a declining function of time).

To incorporate these volatility humps, a very suitable choice is the following parametrization (Rebonato, 1998):

$$\sigma_i(t) = \Phi_i \left(\left[a \left(T_{i-1} - t \right) + d \right] e^{-b(T_{i-1}-t)} + c \right), \tag{4.2}$$

where a, b, c, d, Φ_i are parameters that have to be determined. This is done using a similar procedure as in the previous model: first choose a, b, c, d to find a close fit with swaption and cap prices. Next use Φ_i to ensure σ_i is exactly fitted to caplet prices.

4.4 Correlation

Under their own measure, caplets only depend on the value of a single Libor rate. Therefore, caplets can not be used to calibrate the correlations in the LMM. Swaption cán be used to do this, as can be seen from equation (3.15), Another possibility is to use historical information to obtain estimates of future correlations. Just as for the volatilities, we have to estimate the correlation structure $\rho_{ij}(t)$ for $T_0 \le t \le T_{(i \wedge j)-1}$. Generally it will be assumed that correlations are constant between reset dates. Still, for a model with N different Libor rates, there are $N(N-1)/2$ correlations that have to be determined. To ensure correlations will always be between -1 and $+1$ and to give some more structure to the correlation matrix (denoted by ρ), again a parametrization is used.

A parametrization that is commonly used is:

$$\rho_{ij} = \cos\left(\theta_i - \theta j\right),$$

where θ_i is usually a function of i and some other parameters (Brigo and Mercurio, 2001).

Another possibility is:

$$\begin{aligned}\rho_{ij} &= \prod_{k=i}^{j-1} \rho_{k,k+1} \quad j > i, \\ \rho_{ij} &= \rho_{ji} \qquad\qquad j < i.\end{aligned}$$

This requires specifying the upper diagonal of the correlation matrix, the correlation between subsequent Libors (see Schoenmakers and Coffey, 2003). A special case is when $\rho_{k,k+1} = e^{-\beta}$ for all k:

$$\rho_{ij} = e^{-\beta|i-j|}, \tag{4.3}$$

where β is the only parameter to be determined. This implies that correlations between Libors are constant through time and only depend on the difference between their reset dates.

4.5 Other issues

There are numerous other issues regarding calibration. Take for example the caplet volatilities. Even though in theory this works perfectly, there are some practical issues. Black's formula assumes that the volatility is independent of the moneyness ($L_i(t)/R$) of the caplets. In practice, the volatility is higher for in- and out-of-the-money caplets, the so-called volatility smile. Therefore *the* caplet volatility does not exist and we will have to adjust our parameters to take this into account.

Furthermore there is the risk of overfitting. Suppose we could calibrate our model to every plain-vanilla product available. Then the resulting volatility and correlation structure can become a very irregular function, which is unrealistic. So we have to make a trade-off between a realistic but not perfectly calibrated and an perfectly-fitted but unrealistic model.

For more information we refer to the references at the begin of this chapter.

Part II
Monte Carlo

5 Introduction to MC

Several methods are available for the valuation of financial derivatives. An overview can for example be found in (Hull, 2003; Wilmott, 2006). For some products, there sometimes exist analytical solutions. The most famous one is of course the Black and Scholes (1973) equation, from which most other option valuation formulas are derived. Analytical solutions are usually only available for plain-vanilla products and simple models. In the Libor Market Model we saw that caplets can be valued analytically. This is because its payoff depends only on a single Libor rate.

For more complicated products, no analytical solutions are available and we have to use numerical procedures. A very common method is the binomial tree. Basically, it replaces the continuous distribution of the underlying by a discrete distribution. Another way to find the value of a derivative, is to solve the differential equation for the value of the derivative numerically, for example by a finite difference method.

Both methods discretize the underlying over time. In the Libor Market Model, the underlying is a vector of Libor rates: $L_i(t), i = 1..N$. The size of this vector is typically between 10 and 120 (quarterly resets with 30 year maturities). In the tree and finite difference method, we have to discretize this in every dimension, so the number of grid points grow exponentially with the number of dimensions (the so-called curse of dimensionality). In general, both methods only work well up to 2 or 3 dimensions. Therefore these methods can not be used for pricing in the Libor Market Model.

The last possibility to value financial derivatives is to use Monte Carlo simulation. Because the methods described above cannot be used, this is the method that is generally used in the LMM. This chapter describes the Monte Carlo method in general. The next chapter gives an overview of possible ways to improve its performance. In Chapter 7 this will be applied to the Libor Market Model.

5.1 Definition

Monte Carlo methods have been used since the introduction of the first programmable computers in the 1940s. They can be applied to a wide range of problems. Boyle (1977) was the first to use these methods for derivative pricing. Monte Carlo methods are easy to implement and applicable to almost every valuation problem, which is the main advantage of the method, together with the ability to solve high-dimensional problems.

Suppose we want to compute the expected value V of a function $f(Z)$, depending on a random variable Z, given a probability density $\psi(Z)$:

$$V \equiv \mathbb{E}\left(f(Z)\right) = \int f(z)\psi(z)dz. \tag{5.1}$$

The Monte Carlo method to solve this can be summarized as follows:

1. Establish a procedure to simulate Z from the probability density $\psi(Z)$.
2. Draw a variate z_m
3. compute $f(z_m)$.
4. Repeat step 2 and 3 for $m = 1..M$ and compute the average. This is the Monte Carlo estimate of equation (5.1).

So the Monte Carlo estimate is given by

$$\widetilde{V} = \frac{1}{M} \sum_{m=1}^{M} f(z_m). \tag{5.2}$$

5.2 Properties

The Monte Carlo estimate is unbiased:

$$\mathbb{E}\left(\widetilde{V}\right) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}\left(f(Z_m)\right) = \mathbb{E}\left(f(Z)\right) = V.$$

The variance of $\widetilde{V}$ is given by:

$$\operatorname{var}\left(\widetilde{V}\right) = \frac{1}{M^2} \sum_{m=1}^{M} \operatorname{var}\left(f(Z_m)\right) = \frac{1}{M}\sigma_f^2,$$

where $\sigma_f^2 = \operatorname{var}\left(f\left(Z\right)\right)$. Usually the standard deviation σ_f is unknown, but we can estimate it by the sample variance:

$$s_f^2 = \frac{1}{M-1} \sum_{m=1}^{M} \left(f(z_m) - \frac{1}{M} \sum_{m=1}^{M} f(z_m) \right)^2,$$

which is an unbiased estimate of σ_f^2.

From the Central Limit Theorem it follows that, for $M \to \infty$, $\widetilde{V}$ converges in distribution to the Normal distribution:

$$\widetilde{V} \xrightarrow{d} \mathcal{N}\left(V, \frac{\sigma_f}{\sqrt{M}}\right).$$

Now we can define the standard error of a Monte Carlo simulation by:

$$\mathrm{se} = \frac{s_f}{\sqrt{M}}. \tag{5.3}$$

5.3 Random number generation

For the implementation of the Monte Carlo method we have to be able to draw random variables Z from a probability density function $\psi(Z)$. The usual way to do this is to draw uniform random variables from the interval $(0, 1)$ and transforming the uniform variables to the desired distribution by (numerical) inversion of the probability distribution function. Because computers are totally deterministic, it is not possible to get truly random numbers. However, numerous (deterministic) algorithms exists that almost perfectly replicate true randomness. Because these algorithms are implemented in almost every piece of numerical software and work correctly for (almost) every application, we will not pay attention to this. For more information see (Jäckel, 2002).

6 Variance reduction

The main disadvantage of the Monte Carlo method is that it converges slowly. As can be seen from the definition of the standard error (equation 5.3), it converges at $\mathcal{O}\left(\sqrt{M}\right)$. So to reduce the standard error of the estimate $\widetilde{V}$ by a factor F, we need to increase the number of samples by F^2. Especially in finance, where derivatives have to be priced both quickly and accurately (an error of only a few basispoints can lead to large losses), this is an important disadvantage.

Several techniques are available to reduce the variance of the Monte Carlo estimate. These will be addressed in the following sections. First we define two benchmarks, denoted by κ and θ, to compare different methods with each other.

Suppose that for a MC implementation u each simulation path m gives an unbiased estimate f_m^u with standard deviation s_u. When we use M simulations, The standard error in the estimate of V is given by $\mathrm{se}_u = \frac{s_u}{\sqrt{M_u}}$ (see equation 5.3). From this we define κ as the relative improvement (measured in se^2) of a variance reduction technique u compared to standard Monte Carlo simulation:

$$\kappa = \left(\frac{\mathrm{se}}{\mathrm{se}_u}\right)^2. \tag{6.1}$$

Because the squared standard error scales linear with the number of simulations (see equation 5.3), κ gives the relative number of paths to use to get the same accuracy. For example, suppose that for a given method u there holds $\kappa = 3$. That means that, to get the same accuracy without control variates, one needs 3 times as many paths.

The relative improvement κ does not take into account the extra computation time that is necessary to value the control variate in each path. Of course, when $\kappa = 2$, but method u requires two times as much computation time for each path, there is no improvement. Using more sophisticated methods to reduce variance, usually also implies more computation time for each simulation. We have to make a trade-off between these two. Define the average simulation time for each path by τ for a standard Monte Carlo simulation and by τ_u for variance reduction technique u. Now we can define θ as the time-adjusted improvement factor by

$$\theta \equiv \kappa \frac{\tau}{\tau_u}. \tag{6.2}$$

The higher θ, the faster the convergence of the simulation. For example, if for a variance reduction technique there holds $\theta = 2$, it takes half the time to get the same standard error as without variance reduction. Therefore, this is the variable we would like to maximize. However, because computation time is very implementation dependent, we will also take κ into account.

6.1 Antithetic variates

A very simple, and widely used, technique in financial pricing problems is the method of antithetic variates (or variables) or antithetic sampling (AS). It uses the fact that if Z is standard normally[2] distributed, so is $-Z$. Suppose ψ is the (possibly more-dimensional)

[2] This is not restricted to normal distributions. For example, if U is uniformly distributed on $[0, 1]$, so is $1 - U$. But because for the LMM only normal distributions are relevant, we will focus on these.

Gaussian probability density and z_m is a variate drawn from this distribution. Then we replace the estimate $f(z_m)$ in equation (5.2) by

$$\overline{f}_m = \frac{f_m^+ + f_m^-}{2},$$

where $f_m^+ \equiv f(z_m)$ and $f_m^- \equiv f(-z_m)$. Because we have to evaluate two paths for each simulation m, we will only use half the normal amount of simulations, to get approximately the same computational time. $M/2$ antithetic pairs $\{z_m, -z_m\}$ are more regularly distributed then M independent variables; for example, the mean is always equal to zero, which almost surely does not hold for the latter. This will probably lead to a reduction in variance. More precise, if $f^+ \equiv f(Z_m)$ (and $f^- \equiv f(-Z_m)$) has variance σ_f^2 then the variance of $\overline{f}_m$ is given by

$$\begin{aligned} \overline{\sigma_f}^2 &= \frac{1}{4}\left(\sigma_f^2 + \sigma_f^2 + 2\operatorname{cov}\left(f^+, f^-\right)\right) \\ &= \frac{1}{2}\left(\sigma_f^2 + \operatorname{cov}\left(f^+, f^-\right)\right). \end{aligned}$$

For the antithetic variates to be effective, we need: $\overline{\sigma_f}^2 < 2\sigma_f^2$, because we need two evaluations (f_m^+ and f_m^-) for a single simulation m. So the technique is effective when $\operatorname{cov}\left(f^+, f^-\right) < 0$. This will usually be the case. It is always true when $f(z)$ is a monotonic function in z. This is an important observation, because the products we will be considering are usually monotonic. A proof can be found in the appendix (B). Note that the use of AS does not always have to beneficial. For some exotic products, it could also be the case that there is no negative correlation between the paths (take for example a butterfly option construction, which has a symmetric payoff).

Usually, sampling a pair of antithetic paths is faster than 2 independent paths. First because now we only need to draw z_m once. Second because it may be easier to compute $f(-z_m)$ after $f(z_m)$ because some of the computations will probably be the same. The improvement is very much depending on the implementation, but will generally be relatively small.

6.2 Control Variates

The method of control variates (CV) is based on the principle 'use what you know'. Suppose we know that the function $f(Z)$ (in our case the value of a financial derivative), is very similar to another function $g(Z)$, for which we know its expected value is G. We can incorporate this information by only simulating the difference between the two products $(f(Z_m) - g(Z_m))$. So we replace f_m in equation (5.2) by:

$$\phi_m = f(z_m) - g(z_m) + G$$

where the expected value G is just a constant in the simulation. If $f \equiv f(Z)$ and $g \equiv g(Z)$ are strongly correlated the variance of ϕ will be smaller than the variance of f.

We can generalize this by multiplying g with a constant β, so we get

$$\phi_m = f_m - \beta g_m + \beta G \tag{6.3}$$

where we denote $f_m \equiv f(z_m)$ and $g_m \equiv g(z_m)$ and where β is a parameter that can be chosen freely. The Monte Carlo estimate is given by:

$$\begin{aligned}\widetilde{V}^* &= \frac{1}{M}\sum_{m=1}^{M}\phi_m \\ &= \frac{1}{M}\sum_{m=1}^{M}(f_m - \beta g_m) + \beta G \qquad (6.4)\end{aligned}$$

This is again an unbiased estimator of V, because

$$\mathbb{E}\left[g(Z_m) - G\right] = \mathbb{E}\left[g(Z_m)\right] - G = 0.$$

The variance is given by:

$$\operatorname{var}(\widetilde{V}^*) = \frac{1}{M}\left(\sigma_f^2 + \beta^2\sigma_g^2 - 2\beta\operatorname{cov}(f,g)\right). \qquad (6.5)$$

Minimizing with respect to β gives:

$$\beta^\star = \frac{\operatorname{cov}(f,g)}{\sigma_g^2}, \qquad (6.6)$$

so that equation (6.5) simplifies to:

$$\operatorname{var}(\widetilde{V}^*) = \frac{1}{M}\sigma_f^2\left(1 - \rho_{f,g}^2\right), \qquad (6.7)$$

where $\rho_{f,g}$ is the correlation between f and g. This shows that every choice for g that is correlated with f will lead to variance reduction. The larger the absolute correlation, the larger also the reduction in variance will be. Of course, the challenge is to find a variable g that can be valued relatively easy and is highly correlated with the derivative that we need to value.

6.2.1 Vector of Control Variates

The same analysis can be extended to a vector of control variates $\mathbf{g}$, by replacing β by a row-vector β'.

$$\phi_m = f_m - \beta'\left(\mathbf{g_m} - \mathbf{G}\right).$$

Again, the Monte Carlo estimate is unbiased. Its variance is given by:

$$\operatorname{var}(\phi_m) = \frac{1}{M}\left(\sigma_f^2 + \beta'\operatorname{var}(\mathbf{g})\beta - 2\beta'\operatorname{cov}(f,\mathbf{g})\right).$$

Now $\operatorname{var}(\mathbf{g})$ denotes the covariance matrix of the vector $\mathbf{g}$ and $\operatorname{cov}(f,\mathbf{g})$ is a column vector where element i is the covariance between f and the ith element of the vector $\mathbf{g}$ The minimum variance is attained when β satisfies:

$$\operatorname{var}(\mathbf{g})\beta = \operatorname{cov}(f,\mathbf{g}). \qquad (6.8)$$

Essentially, this the same as using a single control variate, which is a linear combination of the g_i's. In fact we could use an arbitrary function of the control variates, as long as we can compute its expectation.

6.2.2 Estimating β

The remaining task is to compute β. The optimal value is given by equation (6.6) (or 6.8), but σ_g^2 and $\text{cov}(f, g)$ are usually unknown. They can however be estimated by the sample variance and covariance. The correct way to do this, is to use a separate simulation to estimate β. This has a negative impact on the effectiveness of the control variates, because the time won by the increased convergence is partly lost in the estimation of β. In practice, usually the same simulation is used for the estimation of β as for the Monte Carlo simulation. This will lead to a bias, but in general, it is negligible (Jäckel, 2002).

6.3 Using Low-discrepancy sequences

In normal Monte Carlo simulation we use taking random (or pseudo-random) variables z_m. The idea of low-discrepancy sequences is to replace the random numbers by deterministic numbers in such a way that they look like random numbers, but lead to a faster convergence. These numbers are called quasi-random numbers. Unlike pseudo-random numbers, subsequent draws of quasi-random numbers are no longer uncorrelated.

A drawback of the method is that in general it does not work for very high dimensional problems, where each path consists of a large vector of random numbers. This holds in particular for the Libor Market Model. Recent research however suggest that for applications in finance, quasi random numbers still have fast converge, even for (very) high dimensions. Another drawback of the method of quasi-random numbers is that it is not possible to estimate the standard error accurately.

Even though it may be very interesting to look at the use of low-discrepancy sequences for applications in the LMM, we will not investigate it in this thesis. For more information on random number generation, (Jäckel, 2002) is a good start.

6.4 Other techniques

There are several other methods to reduce, which are listed below. For applications in derivative pricing in the Libor Market Model, they are usually not appropriate, mainly because they do not work for high-dimensional problems.

- **Moment Matching**. When simulating a normal random variable, the moments of the sample distribution will not exactly match the normal moments. By adjusting the moments to make the (first finite number of) moments exactly the same as the normal moments, the convergence speed can be improved. Drawbacks are that the improvements are usually relatively small and (especially for high-dimensional problems) the effort to adjust the samples to match the moments is relatively large.
- **Stratified sampling**. The function domain of $f(Z)$, is divided into J subintervals. For each interval j a representative value $\overline{z}_j$ is chosen. For each variate z_m in the interval j, the value $f(z_m)$ is replaced by $f(\overline{z}_j)$. This reduces the amount of computations, because we only have to compute $f(Z)$ once for each interval. However, it also induces a bias. It will only work when the values within an interval can be chosen such that f is nearly constant on each subinterval. Another, related, definition of stratified sampling is to ensure that the variates are regularly spaced of the probability density ψ, for example by taking the same amount of samples in each percentile of the distribution.
- **Latin hypercube sampling.** This is a high-dimensional extension of the latter definition of stratified sampling. For a high-dimensional problem, it is not always

possible to divide the sample variates evenly over the domain, just because the domain is too large. Latin hypercube sampling is a way to divide samples evenly over the domain. Low-discrepancy sequences usually already ensure that variates are evenly spaced over the domain, which make these methods superfluous.

- **Importance sampling.** In some MC simulations, only a small set of paths has a large influence on the value of a product. For example for a deeply out-of-the-money product, only a few paths will lead to non-zero cash flows. By applying a change of measure, it is possible to concentrate on these paths where a lot of simulations are necessary and ignore the paths where 'nothing happens' (for example for deep OTM products, where only few paths lead to payoffs).

For more information on these methods, see (Boyle, 1977; Boyle, Broadie and Glassermann, 1997; Jäckel, 2002; Glasserman, 2003). These methods will not be investigated in this thesis.

7 Monte Carlo implementation

In order to simulate Libor rates to price exotic derivatives, we have to discretize and rewrite the equation that describes the dynamics of the Libor rates. This will be done in the next section. It turns out to be impossible to give an exact solution for the drift, so we have to approximate it, which will be described in section 7.2. After that, we give some examples of the pricing of simple cash flows and in section 7.4 we explain how we can price derivatives with MC.

7.1 Solving the LMM

We would like to have an explicit expression for $L_i(t)$. We start with equation (3.10):

$$dL_i(t) = \mu_i\left(\mathbf{L}(t), t\right) dt + \sigma_i(t) L_i(t) dW_i(t). \tag{7.1}$$

We can apply Ito's lemma to $\ln L_i(t)$, to find:

$$d\ln L_i(t) = \left(\mu_i\left(\mathbf{L}(t), t\right) - \frac{1}{2}\sigma_i^2(t)\right) dt + \sigma_i(t) dW_i(t). \tag{7.2}$$

Integrating over $[T_0, t]$ gives the following solution:

$$L_i(t) = L_i(T_0)\exp\left(\int_{T_0}^{t} \mu_i\left(\mathbf{L}(u), u\right) du - \int_{T_0}^{t} \frac{1}{2}\sigma_i^2(u) du + \int_{T_0}^{t} \sigma_i(u) dW_i(u)\right). \tag{7.3}$$

If we discretize time t by the tenor dates $T_0 < \cdots < T_N$, we get:

$$L_i(T_{k+1}) = L_i(T_k)\exp\left(X_{ik} + Y_{ik} + Z_{ik}\right), \tag{7.4}$$

where X_{ik}, Y_{ik} and Z_{ik} are the three integrals in equation (7.3) over $[T_k, T_{k+1}]$. We discuss these integrals separately now.

For X_{ik}, if we substitute the drift term, there holds:

$$\begin{aligned} X_{ik} &\equiv \int_{T_k}^{T_{k+1}} \mu_i\left(\mathbf{L}(u), u\right) L_i(u) du \\ &= \int_{T_k}^{T_{k+1}} \sigma_i(u) \sum_j \frac{\sigma_j(u)\rho_{ij}(u)\alpha_j L_j(u)}{1 + \alpha_j L_j(u)} du. \end{aligned}$$

The summation depends on the chosen numeraire (see Section 3.2). Because of the term $L_j(u)$ in the the integral, the drift term is stochastic, we cannot evaluate it exactly and we have to approximate it. For the moment, we will estimate $L_j(u)$ by its value at time T_k:

$$L_j(u) \approx L_j(T_k): \quad T_k \leq u < T_{k+1.} \tag{7.5}$$

This gives:

$$\begin{aligned} X_{ik} &\approx \int_{T_k}^{T_{k+1}} \sigma_i(u) \sum_j \frac{\sigma_j(u)\rho_{ij}(u)\alpha_j L_j(T_k)}{1 + \alpha_j L_j(T_k)} du \qquad (7.6) \\ &= \sum_j \nu_j(T_k) \int_{T_k}^{T_{k+1}} \sigma_i(u)\sigma_j(u)\rho_{ij}(u) du, \end{aligned}$$

where $\nu_j(T_k) \equiv \frac{\alpha_j L_j(T_k)}{1+\alpha_j L_j(T_k)}$. Now define the matrix $C(k)$ with elements $C_{ij}(k)$ $(i, j \geq k+2)$ by:[3]

$$C_{ij}(k) \equiv \int_{T_k}^{T_{k+1}} \sigma_i(u)\sigma_j(u)\rho_{ij}(u)du. \tag{7.7}$$

Then equation (7.6) simplifies to:

$$X_{ik} \approx \sum_j \nu_j(T_k)C_{ij}(k).$$

Even though this is not the exact expression for X_{ik}, due to the approximation in equation (7.5), it turns out that this does not lead to large errors. However, there do exist better approximations, which will be discussed in section (7.2).

For the next integral in equation (7.4), Y_{ik}, we can use equation (7.7) to rewrite it as:

$$Y_{ik} = -\frac{1}{2}C_{ii}(k).$$

The last integral is defined by:

$$Z_{ik} = \int_{T_k}^{T_{k+1}} \sigma_i(u)dW_i(u).$$

This is just a normal variable, with expectation, variance and covariance given by

$$\begin{aligned} \mathbb{E}(Z_{ik}) &= 0 \\ \mathbb{E}\left(Z_{ik}^2\right) &= \int_{T_k}^{T_{k+1}} \sigma_i^2(u)du \\ \mathbb{E}(Z_{ik}Z_{jk}) &= \int_{T_k}^{T_{k+1}} \sigma_i(u)\sigma_j(u)\rho_{ij}(u)du. \end{aligned}$$

So the variance-covariance matrix of the vector $Z_k \equiv [Z_{k+2,k}, \cdots, Z_{N,k}]$ is $C(k)$ (equation 7.7):

$$Z_k \sim \mathcal{N}(\mathbf{0}, C(k)).$$

Now define the lower triangular matrix $H(k)$ by the Cholesky decomposition of $C(k)$, which exists because $C(k)$ is symmetric and positive definite (we write $C \equiv C(k)$ and $H \equiv H(k)$):

$$HH' = C.$$

When w_k is a $(N-k-1)$-vector of independent standard normal distributed variables, then there holds that Hw_k has C as covariance matrix, because:

$$\operatorname{cov}(Hw_k) = H \operatorname{cov}(w_k) H' = HH'I = C.$$

So we can write

$$Z_{ik} \equiv [H(k)w_k]_i,$$

where $[\cdot]_i$ denotes the ith element of a vector.

[3]Because $C_{ij}(k)$ is only defined for $i, j \geq k+2$, the first (upper left) element of the matrix is $C_{k+2,k+2}(k)$. So $C(k)$ is a square matrix with $N-k-1$ rows and columns.

Substituting X_{ik}, Y_{ik} and Z_{ik} back into equation (7.4) we get:

$$L_i(T_{k+1}) = L_i(T_k)\exp\left(\sum_j \nu_j(T_k)C_{ij}(k) - \frac{1}{2}C_{ii}(k) + [H(k)w_k]_i\right). \tag{7.8}$$

So if we can compute the covariance matrix $C(k)$, we can simulate the Libors by this equation. To simulate the Libors up to time T_j, we have to draw j vectors of random variables w_k $(k = 1, .., j)$ Define this vector as z_j[4]:

$$z_j \equiv [w_0, ..., w_{j-1}]. \tag{7.9}$$

To simulate all the Libors (for $T_1, ..., T_{N-1}$) we need the vector $z \equiv z_N$.

Two remarks:

- If the correlation matrix (and therefore C) does not have full rank N but rank $D < N$ it is possible to make the following Eigenvalue decomposition (suppressing dependency on k):
$$C = V\Lambda^2 V^\dagger,$$
where $\Lambda(k)$ is a $D \times D$ diagonal matrix with on the diagonal the square roots of the eigenvalues of the matrix $C(k)$ and $V(k)$ is a $M \times D$ matrix containing the eigenvectors of C. Then we can replace H by $\widetilde{H} \equiv V\Lambda$. This can be used if we do not want to use a full-rank model, but assume that the Libors are driven by D uncertainty factors (see also Chapter 3).
- Instead of applying Ito's lemma to equation (7.1), it is also possible to simulate L_i directly by using an Euler discretization:
$$L_i(T_{k+1}) = L_i(T_k) + dL_i(T_k)\,(T_{k+1} - T_k)\,,$$
where $dL_i(T_k)$ is given by equation (7.1). This leads to a discretization error, because it assumes $dL_i(t)$ is constant for $T_k \le t < T_{k+1}$. This method is less accurate than the one described above.

7.2 Drift correction

In the derivation in Section 7.1 we made the assumption that $L_j(u)$ is constant between reset dates to compute the drift integral (equation 7.5 and 7.6). This leads to a discretization error. Several drift approximating methods have been proposed to reduce this error (e.g. Glasserman and Zhao, 2000; Pietersz, Pelsser and van Regenmortel, 2004; Joshi and Stacey, 2006). Here we will discuss a method proposed by Hunter, Jäckel and Joshi (2001), called the predictor-corrector method. We start from the discretized version of equation (7.3). Suppose we have simulated the Libors, denoted by L_i up to time T_k. Then the next Libor follows

$$L_i(T_{k+1}) = L_i(T_k)\exp\left(\int_{T_k}^{T_{k+1}} \mu_i\left(\mathbf{L}(u), u\right)du + \ldots + f(w_k)\right),$$

where the exponent has been simplified for readability and W is a multivariate standard normal variable. Next we assumed $\mathbf{L}(u) = \mathbf{L}(T_k)$ for all $T_k \le u < T_{k+1}$ to be able

[4]The size of the vector z is the sum of the sizes of the w_k's: $\sum_{i=1}^{k}(N - i - 1)$. When we do not use a full-factor model, the size will be smaller.

to compute the integral. For a given realization of w_k, denoted by $\overline{w}_k$, this gives the simulated Libor at T_{k+1}:

$$L_i(T_{k+1}) = L_i(T_k) \exp\left(\int_{T_k}^{T_{k+1}} \mu_i \left(\mathbf{L}(T_k), u\right) du + f(\overline{w}_k)\right). \tag{7.10}$$

So far this is equal to the method proposed in Section 7.1. We will now improve this estimate by using the information of the simulated value of $L_i(T_{k+1})$. Because we also have the value of L_i at T_{k+1} from equation (7.10), it is possible to improve our initial approximation of the function $\mathbf{L}(T_u)$ ($T_k \leq u < T_{k+1}$) by taking:

$$\bar{\mathbf{L}}(u) = \frac{\mathbf{L}(T_k) + \mathbf{L}(T_{k+1})}{2}, \quad T_k \leq u < T_{k+1}.$$

Using this improved estimate, we can correct our previous simulated Libor by substituting this into equation (7.10), using the same realization of the Brownian motion:

$$\begin{aligned} \widehat{L}_i(T_{k+1}) &= L_i(T_k) \exp\left(\int_{T_k}^{T_{k+1}} \mu_i \left(\bar{\mathbf{L}}(T_k), u\right) du + f(\overline{w}_k)\right) \\ &= L_i(T_{k+1}) \exp\left(\int_{T_k}^{T_{k+1}} \frac{1}{2} \left[\mu_i \left(\bar{\mathbf{L}}(T_k), u\right) - \mu_i \left(\bar{\mathbf{L}}(T_k), u\right)\right] du\right). \end{aligned}$$

Hunter et al. (2001) show this predictor-corrector approach outperforms the original (also called log-Euler) method and say it is possible to evolve the model as far as twenty years in one step. Because steps are typically 3, 6 or 12 months, this method will be sufficiently accurate.[5]

7.3 Arbitrage free pricing of cash flows

From the theory of arbitrage-free pricing, the present value $V(t)$ of a cash flow CF_i at time T_i, which is the value of a product, is given by:

$$\frac{V(t)}{B(t)} = \mathbb{E}_B \left[\left. \frac{\mathrm{CF}_i}{B(T_i)} \right| \mathcal{F}_t \right], \tag{7.11}$$

where $B(t)$ is the numeraire and $\mathbb{E}_B$ denotes the expectation under the $B(t)$-measure. I will give three examples to show some applications.

1. First we will use the T_i-forward measure $p(t, T_i)$ as numeraire. Denote the corresponding martingale measure by Q^i. Then $\frac{f(t)}{p(t,T_i)}$ is a martingale under this measure for all security prices $f(t)$. So for $t < \tau < T_i$, there holds:

$$\frac{f(t)}{p(t, T_i)} = \mathbb{E}_i \left[\frac{f(\tau)}{p(\tau, T_i)} \right], \tag{7.12}$$

[5] It is possible to simulate the Libors without going from one tenor date to the next, but jumping to the final maturity date T_N in one step. For this so-called extreme long jump method (Rebonato, 2002), steps of 20 years may occur. However, this method only simulates Libor rates $L_i(T_{i-1})$ and not the forward Libors ($L_i(T_k)$, $k < i - 1$) which makes the method useless for pricing callable Libor exotics. Therefore this will not be discussed here.

where $\mathbb{E}_i$ denotes the expectation under the Q^i-measure. Now set $f(t) = p(t, T_{i-1}) - p(t, T_i)$. Clearly, this is the price of a security. Substituting this into equation (7.12) with $\tau = T_{i-1}$ and using definitions (1.4) and (1.5), it follows that:

$$L_i(t) = \mathbb{E}_i \left[R_i \right],$$

so the forward rate is equal to the expected future spot rate (under the appropriate numeraire).

2. First take $B(t)$ as the numeraire, defined by equation (3.7):

$$B(t) = \frac{p(t, T_{m(t)})}{\prod_{j=1}^{m(t)} p(T_{j-1}, T_j)} = p(t, T_{m(t)}) \prod_{j=1}^{m(t)} (1 + \delta_j L_j(T_{j-1})).$$

Using this as the numeraire, we find the following, using $f(t) = p(t, T_n)$:

$$\frac{p(t, T_n)}{B(t)} = \mathbb{E}_Q \left[\frac{p(\tau, T_n)}{B(\tau)} \right] = \mathbb{E}_Q \left[\frac{p(\tau, T_n)}{p(\tau, T_{m(\tau)}) \prod_{j=1}^{n} (1 + \delta_j L_j(T_{j-1}))} \right].$$

Taking $t = 0$ and $\tau = T_n$, we get:

$$p\left(0, T_n\right) = \mathbb{E}_Q \left[\frac{1}{\prod_{j=1}^{n} (1 + \delta_j L_j(T_{j-1}))} \right]. \tag{7.13}$$

The right hand side can be estimated by simulating the LIBOR rates under the risk neutral measure. So, under the risk neutral measure, the value of a zero coupon bond, maturing at time T_n can be estimated using simulation under the risk neutral measure. On the other hand, the value of $p\left(0, T_n\right)$ is also given by

$$p(0, T_n) = \frac{1}{\prod\limits_{j=1}^{n} (1 + \delta_j L_j(0))},$$

which is known exactly at $t = 0$. Therefore we can use equation (7.13) to test the simulation. For example it can be used to check the accuracy of the drift approximation from section 7.2.

3. Valuing a caplet. From the definition of the Libor rate (equation 1.4) it follows that $p(t, T_i) L_i(t)$ is the price of a security. Assume a caplet with a payoff at time T_i given by:

$$\Pi(T_i) = \alpha_i \left[L_i(T_{i-1}) - R \right]^+.$$

Taking $p(t, T_i)$ as numeraire, the value of the caplet is given by:

$$\text{capl}_i(0) = p(0, T_i) \mathbb{E}_i \left[\alpha_j \left(L_j(T_{j-1}) - R \right)^+ \right].$$

7.4 Monte Carlo pricing of products

In the previous section we saw how to price single cash flows. The Libor exotics we will be valuing, consist of a series of cash flows $\text{CF}_0, \text{CF}_1, ..., \text{CF}_N$ at time $T_0, T_1, ..., T_N$. Because all these cash flows are martingales (under the numeraire), also the sum is a martingale. Therefore we extend equation (7.11) to

$$V = B(T_0)\mathbb{E}^0\left[\sum_{i=0}^{N} \frac{\mathrm{CF}_i}{B(T_i)}\right], \tag{7.14}$$

The cash flows CF_i and numeraire $B(T_i)$ are $\mathcal{F}_i$-measurable. In most cases (see for example all the products from Chapter 1) they are even $\mathcal{F}_{i-1}$-measurable. This holds for all cash flows we will look at. Usually the cash flow CF_i depends on (a subset of) the Libor rates $L_j(T_k)$ up to time T_{i-1} (for all $0 \leq k \leq i-1, k < j \leq N$)[6]. We call the set of these Libors a path $\mathbf{P}_j$, defined by[7]:

$$\mathbf{P}_j = \left[\begin{array}{cccc} \mathbf{L}(\mathbf{T}_0) & \mathbf{L}(\mathbf{T}_1) & \dots & \mathbf{L}(\mathbf{T}_j) \end{array}\right] = \begin{bmatrix} L_1(T_0) & - & & \\ L_2(T_0) & L_2(T_1) & \ddots & \\ & & \ddots & - \\ \vdots & \vdots & & L_{k+1}(T_j) \\ & & & \vdots \\ L_N(T_0) & L_N(T_1) & \cdots & L_N(T_j) \end{bmatrix}.$$

The whole set of the Libors is given by $\mathbf{P} \equiv \mathbf{P}_{N-1}$. There holds $\mathbf{P}_0 \subset \mathbf{P}_1 \subset \cdots \subset \mathbf{P}_{N-1}$. For a given path $\mathbf{P}$.

For a given path $\mathbf{P}$ we can compute the cash flows and numeraires CF_i and $B(T_i)$ for all $1 \leq i \leq N$. Now we can estimate the present value V (equation 7.14) of a product with Monte Carlo:

1. Draw the vector z^m (equation 7.9) from the multi-normal probability density function $\psi(Z)$.
2. Simulate the path $\mathbf{P}^m$, from equation (7.8).
3. Compute the cash flows $\mathrm{CF}_i^m \equiv \mathrm{CF}_i(\mathbf{P}^m)$ and numeraire $B^m(T_i) \equiv B(T_i; \mathbf{P}^m)$ $(1 \leq i \leq N)$
4. Compute the present value

$$V^m = B(T_0)\sum_{i=0}^{N} \frac{\mathrm{CF}_i^m}{B^m(T_i)} \tag{7.15}$$

5. Repeat steps 1-4 for $m = 1..M$ and compute the Monte Carlo estimate $\widetilde{V}$, the average of V^m (equation 5.2).

[6] Even though the cash flows CF_i and numeraire B_i are at least T_i-measurable, we will see they are often even T_{i-1}-measurable. Especially CF_N and B_N can always be written as T_{N-1}-measurable.

[7] A path does only have to contain the Libor rates. For example, in a stochastic volatility model, the volatilities are also part of the path. Generally $\mathbf{P}_j$ is just all the information up to time T_j, so $\mathcal{F}_j$. The products we consider only depend on the Libors, therefore we use $\mathbf{P}_j$.

Part III
Callable Libor Exotics

8 Callable Libor exotics

In the previous chapters we have defined the Libor Market Model and showed how we can use Monte Carlo simulation to price products whose value is depending on the forward Libor rates. The most important group of products we can value in this framework are the Callable Libor Exotics (CLE). For CLE's there are no analytical pricing formulas available, so we need to value them with Monte Carlo. In this chapter we will explain their general characteristics and describe some of them in more detail.

In this thesis we will consider Callable Libor exotics under some simplifying conditions. The reason is that it makes them easier to describe and it simplifies the notation. I will discuss these simplifications in this chapter (usually in footnotes).

8.1 General structure

A callable Libor exotic (CLE) is the right to enter into a Libor exotic at a certain set of dates. Therefore we start with the definition of a Libor exotic.

A Libor Exotic (LE) is a swap contract to exchange a floating rate for a coupon rate at a given set of tenor dates T_i. An example of a Libor exotic is the plain vanilla swap from section 1.5. The holder of the payer swap pays each date a fixed interest rate and receives the Libor rate. The Libor exotic is a generalization of this concept. Equivalent to the swap, the floating rate is usually the Libor rate[8]. The coupon payment however, can have virtually any imaginable structure. Denote the coupon payment at time T_i by C_i. Then the net cash flow to the holder of the payer Libor Exotic at time T_i $(1 \le i \le n)$ is[9]

$$CF_i = \alpha_i K \left(L_i(T_{i-1}) - C_i \right), \tag{8.1}$$

where α_i is again the daycount fraction between T_i and T_{i+1} and K is the notional, see figure 8.1. The cash flows of the receiver Libor Exotic are opposite to these.

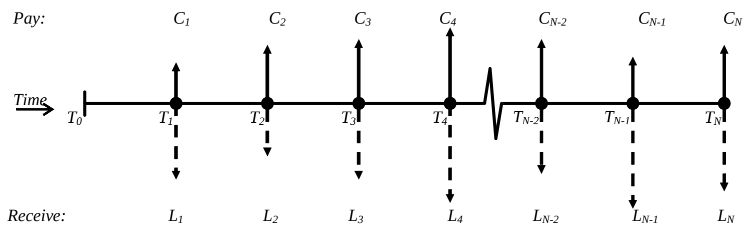


Figure 8.1: Cash flows of a payer Libor exotic

Now we can define the CLE. A Callable Libor Exotic gives the holder the right to enter into the underlying Libor Exotic. Typically, this can be done at (a subset of) the tenor dates[10], which are called the (possible) exercise dates. When the holder of the CLE

[8] Instead of the Libor rate (that pays $L_i(T_{i-1})$ at time T_i), sometimes the Libor-in-arrears rate $L_{i+1}(T_i)$ or even a forward swap rate $R_n^{n+k}(t)$ is used. This will not be discussed in this thesis.

[9] In a more general form, the net cash flow is given by $\alpha_i K \left(g_i F_i - m_i - C_i \right)$, where g is a gearing factor and m is a margin. F_i is not necessarily equal to the Libor rate.

[10] It could as well be possible that the possible exercise dates are are different from the tenor dates. In general, this will not influence on the results of the following chapter, which could equally easy be generalized to an arbitrary set of possible exercise dates $\mathcal{T}_1 < ... < \mathcal{T}_K$. The reason I restrict to tenor dates it that it smplifies notations considerably.

decides to exercise (call) at a certain time τ, the holder receives the underlying Libor exotic and will receive and pay the resulting cash flows (equation 8.1) for all $T_i > \tau$. See figure 8.2.

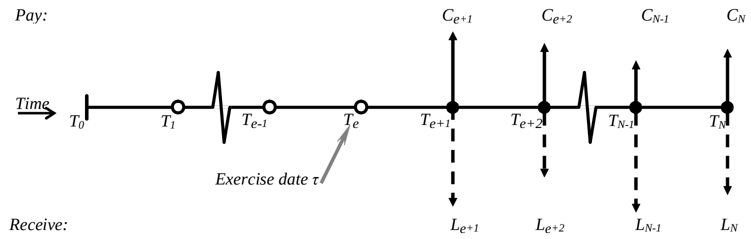


Figure 8.2: cash flows of a CLE when exercised at time $\tau = T_e$

A cancellable Libor exotic (CcLE, also called breakable Libor exotic) is the opposite of the callable Libor exotic. The holder of the CcLE owns a Libor Exotic, but has the right to terminate it on one of the exercise dates. The holder receives the cash flows (equation 8.1) until the CcLE is exercised (cancelled), so for all $T_i \leq \tau$ (figure 8.3).

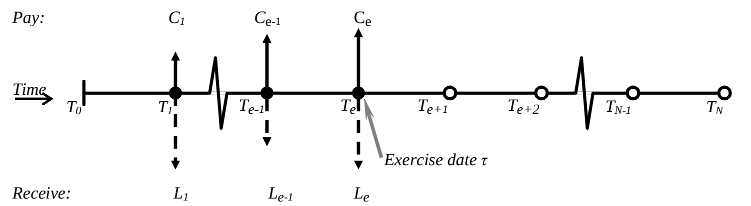


Figure 8.3: cash flows of a CcLE when exercised at time $\tau = T_e$

The choice of the exercise moment τ is made by the holder of the C(c)LE. Of course, he wants to exercise at the optimal exercise moment, which yields him the highest value. The estimation of the optimal exercise date is the subject of the next chapter.

There is a direct relation between (the values of) the cancellable and callable Libor exotics, given by:

$$\text{Cancellable Payer LE} = \text{Callable Receiver LE} + \text{Payer LE}. \tag{8.2}$$

From the definitions, and even more clear from the figures, we can see that this holds whenever the products are exercised at the same time. From this, it immediately follows that the optimal exercise date of both products must be the same. For each exercise date, both sides of equation (8.2) have the same value. Therefore both attain their maximum at the same exercise date, so they have the same optimal exercise date.

8.2 Products

In this section we will describe three callable Libor exotics. The only difference between the different products are the coupon legs C_i of the underlying Libor exotic (equation 8.1). Many more CLE's do exist, but because these will not be used for the results in this thesis, I will not discuss them here.

8.2.1 Bermudan swaptions

The simplest callable Libor exotic is the Bermudan swaption. It's coupon is just a constant:

$$C_i = R_S.$$

In Chapter 1 we already saw the European swaption, which only has a single possible exercise date. Bermudan swaptions are essentially the same, but with more than one exercise date. Contrarily to Europeans, for which there is a very good analytical approximation of the value, there are no analytical pricing methods for Bermudans.

If T_N is the maturity of the underlying swap, a Bermudan swaption can usually be exercised at a set of times T_j $(a \leq j \leq N-1)$. Such a product is called a T_N no-call T_a (also T_N nc T_a, or T_a into T_N) Bermudan swaption.

Bermudan swaptions are by far the most actively traded CLE's and therefore also a lot of research has been published on the valuation of these products (e.g. Andersen, 1999; Jensen and Svenstrup, 2005). Most of the results can be extended to the valuation of other CLE's, but usually this is somewhat more complicated. The reason is that the underlying Libor exotic for a swaption is a swap, which can easily be valued analytically. This does however not hold for most other CLE's.

8.2.2 Callable capped floater

The callable capped floater has a coupon consisting of the Libor rate plus a spread s, capped from above. That means the coupon can never be larger than the cap rate c.

$$C_i = \min\left(L_i + s, c\right).$$

Generally, s and c can be different for each tenor date T_i, in which case we replace them by s_i and c_i.

8.2.3 Callable inverse floater

For a callable inverse floater (CIF) the coupon is the inverse of the Libor rate, capped and floored:

$$C_i = \min\left(\max\left(R - L_i, f\right), c\right),$$

where R is the strike, f is the floor rate and c is the cap rate, all constants. All these could also be different for each cash flow date. In some cases the inverse rate is not capped (so $c = \infty$) and/or floored ($f = -\infty$)

8.2.4 Cancellable Snowball

A snowball is somewhat similar to the callable inverse floater, but now the strike R is no longer constant, but depends on the previous payments. It pays the previous coupon, plus an increment, minus the Libor rate. The coupon payment at time T_i is given by

$$\begin{aligned} C_i &= \min\left(\max\left(C_{i-1} + A_i - L_i, f\right), c\right), \qquad i = 1,..,n, \\ C_0 &= I. \end{aligned} \tag{8.3}$$

Here A_i is the increment and I is the initial previous coupon. From the coupon definition we see that it is path-dependent, since each coupon depends on the previous. If the payoff is not capped, it could grow very large if all Libors L_i are smaller than the increment A_i. If that happens the snowball turns into an avalanche.

9 Valuation of CLE's

To be able to price a callable or cancellable Libor exotic, we have to decide when the product will be exercised. As will be explained in the next section, the value of a CLE can be computed when the exercise date is known. The valuation can be expressed as an optimal stopping problem. .The rest of the chapter shows how to estimate the optimal exercise date.

9.1 Value of a CLE

For simplicity we will assume that the C(c)LE can be exercised at every tenor date $T_1 < ... < T_{N-1}$. Exercising at the last tenor date T_N is useless, because there are no cash flows after T_N. First we look at the CLE. When the callable is exercised at time T_k the holder receives the underlying Libor exotic which leads to cash flows at time $T_{k+1}, ..., T_N$. At time T_k we can compute the value of the CLE, conditional on exercise at time T_k. Discounted to T_0[11], the value is given by (see also equation 7.14):

$$v_k^{call} = B\left(T_0\right)\mathbb{E}^k\left(\sum_{i=k+1}^{N}\frac{\mathrm{CF}_i}{B(T_i)}\right), \tag{9.1}$$

where $\mathbb{E}^k$ is the expectation computed at time T_k, with B as the numeraire. Note that the expectation can sometimes be computed explicitly at time T_k. An example is the case of the Bermudan swaption, where the expectation is simply the value of the swap contract.

For the CcLE, there is one important difference. Suppose we exercise at time T_k.Then the value of the product only consists of cash flows in the past. Therefore, we no longer need the expectation in equation (9.1) and it reduces to:

$$v_k^{canc} = B\left(T_0\right)\sum_{i=1}^{k}\frac{\mathrm{CF}_i}{B(T_i)}. \tag{9.2}$$

The holder of the C(c)LE will exercise the product on the moment that will give him the highest expected value. Otherwise there would exist arbitrage possibilities. So the exercise moment T_k is the moment for which equation (9.1) has its (expected) maximum. That means the value at time T_0 is given by the solution of the following optimal stopping problem (Bender, Kolodko and Schoenmakers, 2006):

$$V = \sup_{\tau\in\mathcal{T}}\mathbb{E}^0\left(v_\tau\right), \tag{9.3}$$

where $\mathcal{T}$ denotes the set of stopping times indices τ: $\mathcal{T} = \{1, ..., N\}$, where $\tau = N$ means the product is not exercised.

To value the product with Monte Carlo, we need to know when the product will be exercised. For each simulation m, starting at T_1, we decide whether we want to exercise, given the information at time T_1 (given by $\mathbf{P}_1^m$). If we do not exercise, we continue with T_{i+1}, until we find the date of exercise. Given this date T_k, the simulated value can be computed from equation (9.1) or (9.2):

$$V^m = v_k(\mathbf{P}_k^m). \tag{9.4}$$

[11] In this and following chapters we will usually discount cash flows at time T_i to their value at T_0. We call this the T_0-value. The reason we do this is to make it possible to compare cash flows occuring at different dates.

Note that this is just a special case of the general pricing of products (equation 7.15). In the case the expectation in equation (9.1) cannot be computed analytically at time T_k, we have to simulate the path until time T_N to get:

$$V^m = B\left(T_0\right) \sum_{i=k+1}^{N} \frac{\mathrm{CF}_i(\mathbf{P}^m)}{B(T_i;\mathbf{P}^m)}.$$

In the following section we discuss how we can estimate the optimal exercise strategy.

9.2 Optimal stopping time

From equation (8.2) we know that the optimal exercise strategy of a callable is equal to the optimal exercise strategy of the opposite cancellable. If we can compute the optimal exercise strategy for one of those, we know it for both. This is easier for the cancellable, because we do not have to compute the expected value of the underlying (compare equations 9.1 and 9.2). Therefore we will try to find the optimal exercise strategy by looking at cancellables only.

First we define $H_m(T_i)$ as the remaining T_0-value of the product, computed at time T_i, in the case we would only have the exercise opportunities T_k for $m \leq k < N$ available $(m \geq i)$:

$$H_m(T_i) = B\left(T_0\right) \sup_{\tau \geq m} \mathbb{E}^i \left(\sum_{j=i+1}^{\tau} \frac{\mathrm{CF}_j}{B(T_j)} \right). \tag{9.5}$$

If we compare this with equation (9.3) together with (9.2) we see that we ignore all the previous cash flow (up to T_i) We call $H_m(T_i)$ the continuation value or hold value of the product. The following properties hold for H:

- the value of the product (equation 9.3) is given by $H_1(T_0)$,
- the value of the CLE at time T_i if we do not exercise is given by $H_{i+1}(T_i)$,
- after the last cash flow, the value is zero: $H_N(T_N) \equiv 0$,
- the value is lower when there are less exercise opportunities $H_i(T_i) \geq H_{i+1}(T_i) \geq \cdots \geq H_{N-1}(T_i)$.
- There holds
$$H_m(T_i) = H_m(T_{i+1}) + \frac{\mathrm{CF}_{i+1}}{B(T_{i+1})} \tag{9.6}$$

Now we can write the determination of the optimal exercise date as follows. Suppose we are standing at time T_j and we have to decide whether we want to exercise (so we did not exercise before T_j). If we exercise (i.e. cancel), we terminate the contract and receive nothing. If we do not exercise, we have the same product, but with one exercise opportunity less, so its value is given by the continuation value $H_{j+1}(T_j)$. From this, it is clear that we will exercise when the continuation value is negative: $H_{j+1}(T_j) < 0$. Therefore, the value of the product at time T_j can be given as:

$$H_j(T_j) = \max\left(H_{j+1}(T_j), 0\right). \tag{9.7}$$

The optimal moment to exercise T_τ the product is thus given by the first date T_j when $H_{j+1}(T_i) < 0$ and T_N if no such date exists:

$$T_\tau = \min\left(T_j : H_{j+1}(T_j) < 0\right) \wedge T_N. \tag{9.8}$$

Now we can define the exercise regions Γ_j. This is the collection of all the situations for which it is optimal to exercise the product:

$$\Gamma_j \equiv \{z \in \Omega : H_{j+1}(T_j; z) < 0)\}\,, \tag{9.9}$$

for each exercise date T_j. From the exercise regions, it is possible to determine the optimal exercise moment for a state:

$$\tau(z) = \min\{j : z \in \Gamma_j\}\,.$$

9.3 Lower bounds

To be able to price a CcLE, we have to compute when the product is exercised for each path m. We first want to know whether it is optimal to exercise at T_1. Therefore we need to know whether $H_2^m(T_1) < 0$. If it is not optimal to exercise at T_1 we continue with the next exercise date and look at $H_3^m(T_2)$.

Because we do not know $H_{i+1}(T_i)$, we have to estimate it. We could estimate $H_2^m(T_1)$ by an inner simulation (for each path). However $H_2^m(T_1)$ depends on $H_3^m(T_2)$, which we also do not know. We would have to use an extra simulation for each inner simulation to estimate it. Continuing in this way we get $N-1$ nested simulation to estimate $H_1, ..., H_{N-1}$. For more than a few exercise dates, this is much to slow.

Clearly, we have to estimate $H_{j+1}(T_j)$ in another way. Several methods have been proposed to do this (e.g. Andersen, 1999; Longstaff and Schwartz, 2001; Kolodko and Schoenmakers, 2006). Every approximation of $H_{j+1}(T_j)$ leads to a suboptimal exercise strategy and therefore the value of the option using this strategy, will be lower than the true option value. The better the estimation of the exercise and continuation values are, the better the exercise strategy and therefore the closer the estimated value is to the true option value. In the next section we will present the method of Longstaff and Schwartz (2001), which is the most widely used method in practice.

9.3.1 Longstaff-Schwartz algorithm

The idea of Longstaff and Schwartz (LS) is to estimate H_i by a function of a set of state variables, where the function is estimated by means of a regression.

For each possible exercise date T_i choose a set of state variables, which are $\mathcal{F}_i$-measurable (so their value is known at T_i). Denote the vector of state variables by $\mathbf{S}$:

$$\mathbf{S}(T_i) = [S_1(T_i, \mathbf{P}_i), S_2(T_i, \mathbf{P}_i), \cdots, S_N(T_i, \mathbf{P}_i)]\,.$$

Furthermore, for each i we choose a parametric family of $\mathbb{R}$-valued functions $f_i(\mathbf{S}; a_i)$. where a_i is a vector of parameters. We will approximate H_i by f:

$$H_{i+1}(T_i) \approx f_i(\mathbf{S}(T_i); a_i) \tag{9.10}$$

Using this approximation, the estimate of Γ_n (9.9) is given by:

$$\widetilde{\Gamma}_n \equiv \{z \in \Omega : f_i(\mathbf{S}(T_i), a_i) < 0\}\,.$$

The parameter vectors a_i have to be chosen in such a way to ensure the best fit in equation (9.10). This is done by optimizing them over a set of K Monte Carlo paths $\mathbf{P}^k \equiv \mathbf{P}_{N-1}(z_k)$. For each path k we can compute the realized cash flows CF_j^k and the corresponding numeraires $B^k(T_j)$ Furthermore, we can compute the state variables for this path for all T_i

$$\mathbf{S}^k(T_i) = [S_1(T_i, \mathbf{P}_i(z_k)), S_2(T_i, \mathbf{P}_i(z_k)), \cdots, S_N(T_i, \mathbf{P}_i(z_k))] .$$

H_i depends on the value of H_{i+1}. To compute a_i we therefore start at time T_{N-1} and work backwards until time T_1.There holds $H_N^k(T_N) \equiv 0$. Now we use the following algorithm:

1. take $n = N - 1$.
2. For all k, compute $H_{n+1}^k(T_n)$ from $H_{n+1}^k(T_{n+1})$ and equation (9.6). Next compute $H_n^k(T_n)$ from equation (9.7):

$$H_n^k(T_n) = \max\left(H_{n+1}^k(T_{n+1}) + \frac{\mathrm{CF}_{n+1}^k}{B^k(T_{n+1})}, 0\right). \tag{9.11}$$

3. Use a cross-sectional regression over all k's of $H_n^k(T_n)$ on $f_n(\mathbf{S}^k(T_i), a_n)$ to estimate a_n.
4. Replace n by $n - 1$ and continue with step 2 until $n = 0$.

In this way, we can compute all parameter vector a_i. This can be used to value the option using Monte Carlo simulation. We can, for each path, compute the state variables $\mathbf{S}(T_i)$ and compute f_i. These can be used to compute the optimal exercise date from equation (9.8). When we know the optimal exercise date, the option is for each path just a series of cash flows, for which we can compute the present value.

See Longstaff and Schwartz (2001) for a simple example of the algorithm for an American put on a non-dividend paying stock.

9.3.2 Implementing LS

To implement LS, we first have to choose what functions f_n we use. Usually f_n is just a polynomial of $\mathbf{S}(T_n)$ with weight vectors a_n. The optimization is then just a linear regression. Under this assumption, the regression can be executed very fast. It turns out that using a polynomial leads a good approximation of the lower bound when we choose good state variables (Piterbarg, 2004). Besides, usually a small set of state variables is enough to get a good exercise strategy. Using to many state variables increases the risk of overfitting.

Usually we do not use all paths in the regressions. For those dates T_i where the next cash flow CF_{i+1} (which is $\mathcal{F}_i$-measurable) is positive, it is never optimal to exercise, because it is always better to cancel at the next exercise date. For callables the same holds when the next cash flow is negative.

In the original article, Longstaff and Schwartz (2001) applied their method to callable derivatives for which the value underlying is known at the moment of exercise. We could do the same for Bermudan swaptions In that case we would have to compare the continuation value with the exercise value E (the value of the swap). It is optimal to exercise as soon as $H_i^m < E_i^m$. Piterbarg (2004) suggests that, when the exercise value is not known, we can use another regression to estimate it. Because this is not necessary when we write the problem in terms of cancellables, the method we presented above is easier and probably better, because we only estimate one variable instead of two.

Joshi and Kainth (2005) propose to use a non-parametric regression. The main advantage is that the risk of overfitting is lower. However the implementation is less straightforward and it is much slower.

9.4 Upper bounds

Every method that gives an estimate of the optimal exercise boundaries (section 9.3) gives a lower bound on the value of the C(c)LE. It does not give an indication of whether the lower bound is close to the true price. For LS, trying different state variables and parametric functions can give an indication of whether it is possible to improve the lower bound. But still it could be possible that we are far from the optimal value.

Some methods have been suggested to obtain an upper bound on the value of the CLE (Rogers, 2002; Haugh and Kogan, 2004). It is not hard to devise an upper bound. At the end of each simulated path (at time T_N), we can look back and determine what the optimal exercise strategy would have been. This can be done by computing the value of the CLE conditional on each exercise date and taking the maximum. Starting from equation (9.3) we get:

$$\begin{aligned} V &\equiv \sup_{1 \leq k < N} \mathbb{E}^0 (v_k) \\ &\leq \mathbb{E}^0 \left(\max_{1 \leq k < N} v_k \right). \end{aligned} \tag{9.12}$$

One can doubt whether this leads to a very close upper bound. It depends on whether the exercise strategy with hindsight is close to the exercise strategy without. The value of the upperbound is equal to the true option value when equation (9.12) holds with equality.

Andersen and Broadie (2004) present an extension to this method, by saying that we can add and subtract an arbitrary martingale to equation (9.12). Let π_i be a martingale for which $\sup_{1 \leq k < N} \pi_k < \infty$. Equivalent to equation (9.12) we can write:

$$\begin{aligned} V &= \sup_{1 \leq k < N} \mathbb{E}^0 (v_k + \pi_k - \pi_k) \qquad (9.13) \\ &= \pi_0 + \sup_{1 \leq k < N} \mathbb{E}^0 (v_k - \pi_k) \\ &\leq \pi_0 + \mathbb{E}^0 \left(\max_{1 \leq k < N} (v_k - \pi_k) \right). \end{aligned}$$

Because this holds for all martingales π we can take the infimum over all martingales Π:

$$V \leq \pi_0 + \inf_{\pi \in \Pi} \mathbb{E}^0 \left(\max_{1 \leq k < N} v_k - \pi_k \right). \tag{9.14}$$

Now have a look at the function $H_j(T_j)$ from equation (9.5). It is a supermartingale ($H_j(T_j) \geq \mathbb{E}^j (H_k(T_k))$, for all $k > j$) so there exists a Doob-Meyer decomposition of the form

$$H_j(T_j) = M_j - A_j, \tag{9.15}$$

where M_j is a martingale and A_j is an increasing process with $A_0 = 0$, so $M_0 = H_1(T_0)$. Now we take $\pi_j = M_j$. Then

$$\begin{aligned} V &\leq M_0 + \mathbb{E}^0 \left(\max_{1 \leq k < N} (v_k - M_k) \right) \\ &= H_1(T_0) + \mathbb{E}^0 \left(\max_{1 \leq k < N} (v_k - H_k(T_k) - A_k) \right) \\ &\leq H_1(T_0) = V, \end{aligned}$$

where we use that $V = H_1(T_0)$ The expectation is negative since $v_k \leq H_k(T_k)$ (see equation 9.11) and $A_k \geq 0$. So for $\pi_j = M_j$ equation (9.13) holds with equality. Because we do not know $H_j(T_j)$ in equation (9.15), we will use an approximation. Because this will affect π_j the equality no longer holds and we find an upperbound from equation (9.14). The better the approximation of $H_j(T_j)$ is, the better the upperbound will be. We already have found an approximation of $H_j(T_j)$ by the lower bound approximation. Andersen and Broadie (2004) describe how to estimate M_j from $\widetilde{H}_j$ an inner simulation, which gives us an upper bound.

Because we need an extra simulation, the computation of an upper bound is very time consuming. However, it is not necessary to compute it every time. For example, when we have chosen a set of state variables and parametrizations in the LS-framework, we can compute an upper bound to test whether the lower bound is close enough to the true value. Once we are content with the result, we can use the lower bound as an approximation, without computing an upper bound each time.

For the results in this thesis, no upper bounds have been computed. We have tested whether the computed lower bounds were correct (and thus whether we have chosen the right state variables and parametrization) by comparing the results to known upper bounds from literature (e.g. Bender et al., 2006).

10 Rasmussen

In the next part, we will apply the use of control variates to the valuation of CLE's. Because the value of the control variate can be computed at every tenor date, we can choose at which date we do this. Rasmussen (2005) shows what the optimal valuation date is for American options for which the value of underlying can be computed analytically. In this chapter we will give his results when applied to the valuation of CLE's. The next section states the problem and section 10.2 gives and proves the optimal valuation date.

10.1 Evaluation moment

For each simulation the value of a CLE is given by equation (9.4), where v_k is given by equations (9.1) or (9.2). The value is computed at the exercise date T_e unless the underlying can not be valued analytical (equation 9.1) and we have to simulate until the end (T_{N-1}) before we can value it.

Now look at the control variates. The control variate is defined by a set of cash flows CF_i Its value at T_0 is given by

$$\Phi = B(T_0)\mathbb{E}^0\left(\sum_{i=1}^{N}\frac{\mathrm{CF}_i}{B(T_i)}\right).$$

For all control variates, this value can be computed analytical at T_0. To use it as control variate, we also value it by Monte Carlo simulation. For simulation m the value is given by equation (7.15). However, in the MC simulation, it is not necessary to simulate the product up to time T_N, since we can price it at every moment T_j. Because the cash flows can be valued analytically, we can define the simulated value, computed at time T_j as:

$$\begin{aligned}
\Phi_j^m &\equiv B(T_0)\mathbb{E}^j\left(\sum_{i=1}^{N}\frac{\mathrm{CF}_i}{B(T_i)}\right)\\
&= B(T_0)\left(\sum_{i=1}^{j}\frac{\mathrm{CF}_i}{B(T_i)}+\mathbb{E}^j\left(\sum_{i=j+1}^{N}\frac{\mathrm{CF}_i}{B(T_i)}\right)\right).
\end{aligned}$$

Note that $\Phi \equiv \Phi_0^m$, the analytical value. Because the value of the control variate is a martingale, there holds for all j:

$$\Phi = \mathbb{E}^0\left(\Phi^m\right) = \mathbb{E}^0\left(\Phi_j^m\right).$$

So in the MC simulation, we can value the product at the moment of our choice. In the following section we show what the optimal valuation date is.

10.2 Optimal evaluation moment

Suppose we want to value a product, which can be valued at a (possibly stochastic) time T_e, whose value is V_e. We use Φ as control variate, which we can value at any time T_i, as explained in the previous section. We want to choose T_i such that the correlation between V_e and Φ_i is as high as possible. Equivalently, we want the variance of $V_e - \beta\Phi_i$ to be as small as possible, where β is between -1 and $+1$. Suppose V_e and Φ_i are positively correlated (so $\beta > 0$) and take for simplicity $\beta = 1$. Then we want to minimize the variance of (see equation 6.3):

$$V_e - \Phi_i = (V_e - V_i) + (V_i - \Phi_i).$$

From the right hand sight we can see the variance can be decomposed into two parts. The first is the variance of the difference in value of the product between the valuation moment of the product and the valuation moment of the control variate. It can be expected to be an increasing function of $|e - i|$. The second term depends on the difference between the value of the control variate and the product at the moment of exercise of the control variate. In general, it will be increasing in i. From this, one could expect that taking $i = e$ is always better than taking $i > e$. Rasmussen (2005) shows this is true. We present a proof below.

We want to show that for all $i > e$, there holds:

$$|\rho(V_e, \Phi_e)| \geq |\rho(V_e, \Phi_i)|.$$

This is the same as showing:

$$\frac{|\mathrm{cov}(V_e, \Phi_e)|}{\sqrt{\mathrm{var}(V_e)\,\mathrm{var}(\Phi_e)}} \geq \frac{|\mathrm{cov}(V_e, \Phi_i)|}{\sqrt{\mathrm{var}(V_e)\,\mathrm{var}(\Phi_i)}}.$$

Because we use the absolute value of the correlation, the statement also holds when the product and CV have a negative correlation. First, we look at the covariances:

$$\begin{aligned}
\mathrm{cov}(V_e, \Phi_i) &= \mathbb{E}^0(V_e\Phi_i) - \mathbb{E}^0(V_e)\,\mathbb{E}^0(\Phi_i) \\
&= \mathbb{E}^0(\mathbb{E}^e(V_e\Phi_i)) - \mathbb{E}^0(V_e)\,\mathbb{E}^0(\mathbb{E}^e(\Phi_i)) \\
&= \mathbb{E}^0(V_e\Phi_e) - \mathbb{E}^0(V_e)\,\mathbb{E}^0(\Phi_e) \\
&= \mathrm{cov}(V_e, \Phi_e),
\end{aligned}$$

where we use the optional sampling theorem, which states that for all Martingales Φ there holds $\Phi_e \equiv \mathbb{E}^e(\Phi_e) = \mathbb{E}^e(\Phi_i)$ for all $i > e$. For the variance, there holds

$$\begin{aligned}
\mathrm{var}(\Phi_i) &= \mathbb{E}^0\left(\Phi_i^2\right) - \mathbb{E}^0(\Phi_i)^2 \\
&= \mathbb{E}^0\left(\mathbb{E}^e\left(\Phi_i^2 - \Phi_e^2 + \Phi_e^2\right)\right) - \mathbb{E}^0(\mathbb{E}^e(\Phi_i))^2 \\
&= \mathbb{E}^0\left(\mathbb{E}^e\left(\Phi_i^2 - \Phi_e^2 + \Phi_e^2\right)\right) - \mathbb{E}^0(\mathbb{E}^e(\Phi_i))^2 \\
&= \mathbb{E}^0\left(\mathbb{E}^e\left(\Phi_i^2\right) - \mathbb{E}^e(\Phi_i)^2\right) + \mathbb{E}^0\left(\Phi_e^2\right) - \mathbb{E}^0(\Phi_e)^2 \\
&= \mathbb{E}^0(\mathrm{var}^e(\Phi_i)) + \mathrm{var}(\Phi_e) \\
&\geq \mathrm{var}(\Phi_e).
\end{aligned}$$

From this it follows that $|\rho(V_e, \Phi_e)| \geq |\rho(V_e, \Phi_i)|$, so we are finished.

We can conclude it is never optimal to value the control variate later than the valuation date of the C(c)LE. For callables where the underlying is analytical and for all cancellables this is the moment of exercise of the product. When the underlying product of the callable cannot be valued analytically, this does not hold (see also section 9.1 on the Monte Carlo valuation of C(c)LEs). In section 12.3 an example will be given which shows the effectiveness of Rasmussen's idea.

We also expect it will not be optimal to exercise at a date $i < e$, because we are ignoring the information between T_i and T_e for the valuation of Φ, while we do use it for V. In section 12.3 we will also give an example of this. Even though it holds in general, we cannot proof that $i < e$ is not optimal. It is not hard to think of a counter example (take a control variate where payoffs between T_i and T_e which are uncorrelated with V).

Part IV
Results

11 Setup

In the previous chapters we have described all the essential ingredients for the valuation of CLE's in the Libor Market model. We have explained the LMM, CLE's, Monte Carlo and showed how we can value a CLE with Monte Carlo in the LMM. Before we can apply this theory to actually price CLE's, we have to choose model parameters and specify the products we want to value. This will be done in the following sections.

11.1 Objectives

In the following chapters, we will try to reduce the standard error of the Monte Carlo simulation. What we would like to achieve is a generic method that can be used to reduce the variance of the simulated price for a range of different product under different (market) circumstances. We are not looking for a control variate that only works for a specific case, because it has no use in practice.

To judge whether a method is effective, we have to choose benchmarks. These will be the relative variance reduction κ and its time-adjusted version θ, as defined in Chapter 6. The latter is the actual improvement in time, but since this can be very implementation dependent we will also consider κ.

For the determination of the time used, we will only take the simulation time into account. So we exclude the time needed to compute the exercise strategy by the LS-method. In practice, it is almost always necessary to compute the exercise strategy, which increases the computation time and decreases the relative effectiveness of variance reduction techniques. Within LS, the time to estimate the optimal exercise strategy depends on the choice of state variables and the number of simulations. Because these choices all influence the simulation time, we decided not to take it into account.

11.2 Model parameters

As explained in Chapter 4, we have to calibrate our model to actual market parameters. In this thesis, we will use the same parameters as has been used in Bender, Kolodko and Schoenmakers (2005). That means we will take tenor dates

$$T_i = 0.5i \qquad (i = 0, .., 12 \equiv N),$$

so the final maturity is $T_N = 6$ and the last Libor is $L_{12}(t) = F(t, 5.5, 6)$. The daycount fraction simply is assumed to be $\alpha_i = T_i - T_{i-1} = 0.5$. We will assume the following time-homogeneous volatility structure:

$$\sigma_i(t) = \Phi_i \left(\left[a\left(T_{i-1} - t\right) + d \right] e^{-b(T_{i-1}-t)} + c \right),$$

as has been proposed in section 4.3.2 (equation 4.2). Furthermore, the correlation between Libor rates is given by (see equation 4.3, section 4.4):

$$\rho_{ij} = \exp\left[\frac{|j-i|}{N-2} \ln \rho_\infty \right], \qquad 2 \leq i, j \leq N.$$

For this choice of parametrization it is possible to compute the covariance matrix (equation 7.7) can be computed analytically. The values of the parameters $\Phi_i, a, b, c, d, \rho_\infty$ and the initial Libors $L_1(0), ..L_N(0)$ are the same as in Bender et al. (2005), and can be found there and in appendix C. See figure 11.1 for a plot of the initial forward term structure,

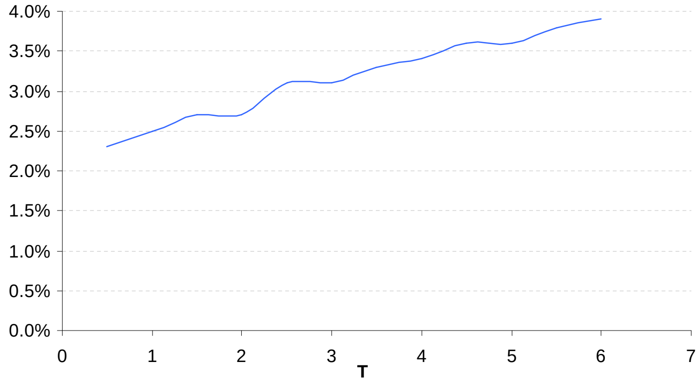


Figure 11.1: Initial term structure $L_i(0)$.

which is upward sloping. Figure 11.2 gives the forward volatility structure at different moments t. As we can see, the volatility is hump-shaped as discussed in section 4.3.2, except for the initial volatility structure ($t = 0$).

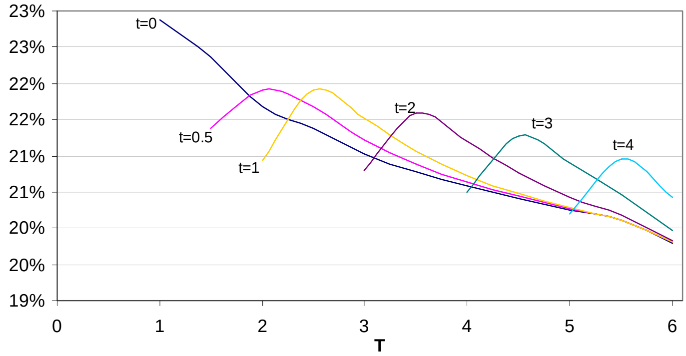


Figure 11.2: volatilities $\sigma_i(t)$ for $t = 0, 0.5, 1, 2, 3, 4$

We will use a full-factor model. As numeraire we take the risk neutral bank account (equation 3.7). For the simulation of Libors we have used the predictor-corrector drift-correction from section 7.2.

11.3 Products characteristics

We will value several products, a Bermudan swaption, a cancellable swap, a callable inverse floater and a cancellable snowball. We will look at ATM, ITM and OTM products. These will be described in the next sections. For all products we will set the principal equal to 1: $K = 1$.

11.3.1 Bermudan Swaption

We will value a T_{12} no-call T_1 Bermudan Swaption (see section 8.2.1) on a payer swap with fixed maturity T_{12}. That means the option can be exercised at time $T_1, .., T_{11}$. When the option is exercised at time T_e the holder receives a $T_e \times (T_{12} - T_e)$ fixed maturity payer swap that pays him at each time T_i, $(e + 1 \leq i \leq N)$ a (possibly negative) amount of (using $L_i \equiv L_i(T_{i-1})$):

$$\alpha_i \left(L_i - R_S\right). \tag{11.1}$$

To make sure the underlying product is at the money (ATM), R_S is set equal to the initial swap rate R_0^N, which assures the value of the swap is 0 at time 0 (see also section 1.5 on swaps). Given the parameters from section 11.2, there holds: $R_0^N = 0.0322$. For the ITM/OTM case, we decrease/increase the strike rate by 1 percent point: $R_S^{ITM} = 0.0222$, $R_S^{OTM} = 0.0422$.

We will also look at the receiver swaption, for which equation (11.1) is replaced by

$$\alpha_i \left(R_S - L_i\right).$$

The ATM strike is the same as for the payer swaption. For the ITM/OTM case, the fixed rates are interchanged: $R_S^{ITM} = 0.0422$, $R_S^{OTM} = 0.0222$.

Finally, we will also look at the corresponding cancellables. These have the same underlying swap, exercise dates and strikes as the callables.

11.3.2 Callable inverse floater

Now we take a callable inverse floater (see section 8.2.3) with the same possible exercise dates $(T_1, ..., T_{11})$ and the same cash flow dates $(T_2, .., T_{12})$ for the underlying product as for the Bermudan swaption. So, when exercised at time T_e, the holder of the product receives the following cash flows at time T_i $(e + 1 \leq i \leq N)$

$$\alpha_i \left(L_i - \min\left(\max\left(R_i - L_i, f_i\right), c_i\right)\right),$$

We will assume $c_i = \infty$[12], $f_i = f$ and $R_i = R$ for all i, so we can rewrite this as:

$$\alpha_i \left(L_i - f - \left[R - f - L_i\right]^+\right), \tag{11.2}$$

where $[\cdot]^+ \equiv \max(\cdot, 0)$. This is just the sum of a Libor rate, a constant and a floorlet. So, just as the underlying of the Bermudan swaption (i.e. the swap), also the underlying of the CIF can be valued analytically.

We take $R = 0.062243$ and $f = 0.02$. With these parameters, the value of the underlying is equal to zero (<0.02 bp) so the product is ATM. For the ITM/OTM case we in-/decrease R again with one percent point.

11.3.3 Cancellable snowball

To test our results on a more complicated product, we will also produce results for the cancellable snowball (see section 8.2.4). The product can be cancelled at dates $T_1, ..., T_{11}$ and will produce cash flows from T_2 until the cancellation date and until T_{12} if it is not cancelled at all. We will use a somewhat simplified version of the general snowball payoff

[12] The same analysis can be performed with a different c. However, this will lead to more complicated payoff functions.

(equation 8.3), which is no longer capped and is floored at zero ($c = \infty, f = 0$). The cash flows at time T_i ($2 \leq i \leq e$, when cancelled at time T_e), are given by:

$$\alpha_i \left(L_i - C_i\right),$$

where the coupon is given by (taking $A_i = A$ for all i):

$$C_i = \left[C_{i-1} + A - L_i\right]^+ .$$

We will take $A = 0.031$ and $C_1 = 0.0135$. These parameters are again chosen to ensure the product is at-the-money (<1bp), where the value of underlying contract (so without the possibility of cancellation), has been computed by simulation, since no analytical solution exists.

We investigate the cancellable instead of the callable, since it is traded more actively on the market. Even though it can be applied just as easy, the callable has a somewhat strange profile, because after calling the next cash flow depends on the previous coupon, which has never been paid. For the callable we therefore would have to keep track of the coupon of the underlying, even if it has not been called yet.

11.4 Longstaff-Schwartz implementation

Before we can start valuing the products described in the previous section, we have to estimate the optimal exercise strategy. We do this by using the Longstaff-Schwartz algorithm, as described in section 9.3.1. To estimate the continuation value H_i, we will use the following state variables:

$$\begin{aligned}
S_1(T_i) &= 1 \\
S_2(T_i) &= \sum_{j=i+1}^{N} \alpha_j p(T_i, T_j) \\
S_3(T_i) &= \sum_{j=i+1}^{N} \alpha_j L_j(T_i) p(T_i, T_j) \\
S_4(T_i) &= \sum_{j=i+1}^{N} \alpha_j L_j(T_i) p(T_i, T_j) \left(i - i_{mid}\right) \\
S_5(T_i) &= \mathrm{CF}_{i+1},
\end{aligned}$$

where $i_{mid} = \frac{i+N}{2}$. S_2 is equal to the present value of a basis point (see equation 1.7), so it corresponds to a vertical shift in the yield curve. S_3 is called the present value of a floating leg. S_4 is called the present value of a curve tilt and corresponds to a relative shift between the Libors with a short time to maturity compared to the Libors with maturities further away. Finally $S_5(T_i)$ is the next cash flow of the underlying Libor exotic (equation 8.1). This is possible, because the cash flow at T_{i+1} is $\mathcal{F}_i$-measurable for all derivatives we will be looking at (If this would not be possible, we should replace CF_{i+1} by CF_i.). Now we will estimate the continuation value H_n by a linear function in the state variables and some quadratic terms of the state variables:

$$f_i(\mathbf{S}(T_i); \alpha_i) = \sum_{j=1}^{5} a_i^j S_i(T_i) + a_i^6 S_3(T_i)^2 + a_i^7 S_4(T_i)^2 + a_i^8 S_4(T_i) S_4(T_i), \tag{11.3}$$

where the weights a_i^j are determined by regression. For the snowball, which the next cash flow CF_{i+1} has more impact, we also add $a_i^9 S_5(T_i)^2$.

As argued in section 9.3.2, we only take those paths into account where the next cash flow CF_{i+1} is negative. When the next cash flow is positive, it can never be optimal to exercise, because it is always better to wait until the next exercise date.

To estimate the factors a_i we follow the procedure described in section 9.3.1. Because equation (11.3) is just linear in the parameters a_i^j, we can simply use ordinary least squares (OLS) to estimate the factors. 50 000 paths have been used. We did not use the same paths for the LS algorithm as for the valuation of the products.

12 Bermudan swaption

In this chapter we look at the valuation of Bermudan swaptions, as described in section 11.3.1. Because the Bermudan swaption is the most simple CLE, it is a good place to start. We will look at several ways to increase the convergence speed of the Monte Carlo simulation by using control variates. In Chapter 14 we will try to use the same ideas for other products.

We start with the straightforward (i.e. without CV's) valuation of the Bermudan swaption in the next section. Then we will discuss the choice of suitable control variates. Next, in section 12.3, we give an example of the use of control variates and illustrate the idea of Rasmussen (2005) as explained in Chapter 10.

In section 12.4 until 12.6 we present results, all for the ATM Bermudan payer swaption. Successively, we look at different single control variates, using a vector of control variates and the use of antithetic sampling. Section 12.8 looks at other Bermudan swaptions: receivers, ITM or OTM swaptions and cancellables, to see whether the results from the previous sections are consistent for all types of swaptions.

A similar analysis has been done by Jensen and Svenstrup (2005) for the Bermudan payer swaption. We will compare the results in the last section.

12.1 Product values (benchmark)

All values and standard deviations in this thesis will be denoted in basis points (bp), which is 0.0001, or $\frac{1}{100}$ of 1% (of the principal K). We use 100000 simulations to estimate the value of the products, without the use of any variance reduction technique.

	Payer		Receiver	
	$\widetilde{V}$	se	$\widetilde{V}$	se
ATM	224.2	0.77	139.3	0.44
ITM	514.1	0.76	508.0	0.68
OTM	100.6	0.57	15.64	0.13

Table 12.1: Bermudan swaption values and standard erros (in bp)

Table 12.1 gives the estimated value ($\widetilde{V}$) and the standard error (se) of different Bermudan swaption. Of course, there holds $\widetilde{V}_{\mathrm{ITM}} > \widetilde{V}_{\mathrm{ATM}} > \widetilde{V}_{\mathrm{OTM}}$. Another, less obvious, observation is that $\widetilde{V}_{\mathrm{Payer}} > \widetilde{V}_{\mathrm{Receiver}}$. This can be explained by the term structure, which is upward sloping. For the ATM swaption, at T_0 the fixed rate R_S is chosen such that the underlying swap is ATM. Because the term structure is upward sloping, the payer swap generally becomes ITM, while the receiver swap becomes OTM. Therefore we can expect the payer swaption to have a higher value. This holds to a lesser extend for ITM swaptions, since both the payer as well as receiver swaption are ITM and will be exercised early.

12.2 Choice of control variates

The aim of this chapter is to find suitable control variates to reduce the computation time for the pricing of Bermudan swaptions. There is a large pool of possible control variates available to choose from. Everything can be used, as long as it has an analytical value. As shown in section 6.2 (equation 6.7) the best control variate is the one with the highest correlation with the Bermudan. We will restrict our attention to products that can be valued within the LMM. Because these products depend on the same quantities (the Libor rates) we can expect to be able to get high correlations.[13] Examples are the products from Chapter 1: forwards, swaps, caplets, floorlets, caps, floors. However, it is not necessary to use existing products: every control variate can be used as long as the expected value is known. Furthermore, all possible linear combinations of control variates are again control variates.

For the Bermudan swaption a few control variates seem to be a natural choice. First one is the (underlying) swap. When the Bermudan is exercised, its value will be equal to the remaining value of the swap. Before exercising, the swap does also lead to some payoffs that will not occur for the Bermudan. Because these will probably be negative (otherwise it would have probably been optimal to exercise the Bermudan earlier), the swap value will be lower than the value of the Bermudan. The more in-the-money the Bermudan is, the earlier it will probably be exercised and the more effective the swap will be as a CV.

Another obvious control variate is the European swaption. Each European is a lower bound for the value of the Bermudan, and they will only be perfectly correlated for those paths where the exercise date of the Bermudan is equal to the European one. Because European swaptions can not be valued analytically in the LMM, they would not be available as control variates. However, because the approximation from equation (1.12) with (3.15) is very accurate, one can argue they can still be used. Although small, this will lead to a bias in the value.

A bit less trivial choice is the cap, but the payoffs of the cap and the Bermudan are very similar. Take a cap with the same cash flow dates as the swap underlying the Bermudan and with a strike of R, close to, or equal to the fixed rate R_S used in the Bermudan. The cap has the same payoffs as the swap, whenever they are non-negative, so it is not hard to see it provides an upper bound for the value of the Bermudan. We will come back to this later on.

The control variate technique will be applied in the following way. For each Monte Carlo simulation m we simulate a path of Libor $\mathbf{P}^m$. From this path, we can compute the exercise moment and value the CLE. We value the control variate(s) on the same path. From all M simulations, we compute the sample covariance between the CLE and the CV as well as the sample variance of the CV. From this we estimate β from equation (6.6). After that, we can compute the estimate of the product value and standard error from equations (6.4) and (6.5). So we use the same simulation for the estimation of β as for the valuation, as discussed in section 6.2.2.

12.3 When to value the control variate

As discussed in Chapter 10, Rasmussen (2005) showed it is never optimal to value the control variate later than the CLE. This will be illustrated in this section, by valuing the ATM Bermudan payer swaption.

For each simulation,the value of the Bermudan will be computed at the exercise moment T_e (so we have a cash flow equal to the swap value at that moment). Note that is not possible to value before the exercise moment, because the moment of exercise is not

[13] Piterbarg (2004) proposes a total different approach. He suggests to use a one or two-factor model and take the value of the CLE in this model as a control variate. The 'analytical' solution is obtained from a PDE method (which is possible in a model with only a few factors). In a few-factor model, not all correlations and volatilities can be taken into account, but we can expect that the value is a reasonable approximation of the true value ot the CLE. If we simulate the value of the CLE in this model with the same noise terms as in the LMM, we can expect to get a high correlation with value in the LMM.

known in advance (it is an adapted stopping time). It is possible to value the Bermudan swaption after the exercise date. In that case the standard error of the simulated value will become larger, due to the uncertainty in the swap value[14]. and therefore we can expect this is worse than valuing at T_e.

As a control variate we take a cap with the same cash flow dates as the Bermudan $(T_2, .., T_{12})$ and also the same strike $(R = 0.322)$. The control variate can be valued at any time: If T_e is the exercise date, all caplets maturing before or at T_{e+1} have known cash flows. All other caplets are valued using Black's formula (equation 1.10), as explained in section 10.1. Because we can choose the valuation date freely, we have to look for the moment with the highest correlation with the CLE. Rasmussen (2005) showed valuing at T_e is always better than valuing at $T_i > T_e$. We will illustrate this by comparing the situations with valuation at T_e, at T_{e+1} and valuation at T_{11} (so when all cash flows are known). Because there is no such proof for valuation before T_e we will test it empirically. It is very tempting to take for example T_{e-1} as the valuation date, but this is not possible. The reason is that the valuation date T_i has to be adapted to $\mathcal{F}_i$. At time T_{e-1} we do not yet know that we are going to exercise at time T_e, so how would we know that we have to value the control variate at that time? Doing so would result in a incorrect price for the control variate[15]. Therefore we will exercise it at

$$\min\left(T_e, T_u\right),$$

where u is fixed for all simulations. In this way we are sure that the valuation date is smaller than or equal to the exercise date of the CLE. We will take $u = 3$, $u = 6$ and $u = 9$. 100000 simulations are used. The resulting κ's are shown in figure 12.1. θ is not given, because these are hardly different from the κ's. Please note the y-axis has a logarithmic scale.

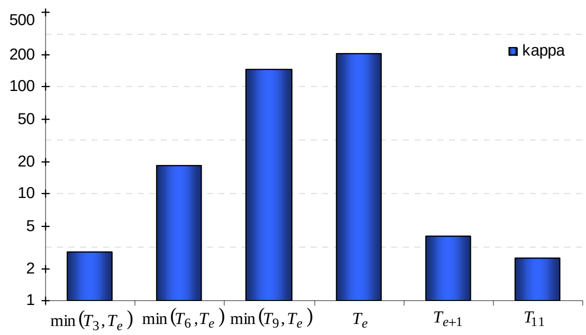


Figure 12.1: κ for the cap as control variate for the Bermudan payer swaption, for different valuation dates for the cap.

We see that valuing at T_e is indeed optimal. From last three bars we see it is (much) worse to valuate after the exercise date. When the CV is valuated at T_{11} the relative improvement κ is 2.5. That means that, to get the same accuracy we only need $\frac{1}{2.5}$ times as many paths as without control variates. So with $\frac{100000}{2.5} = 40000$ simulations we get the same accuracy as with 100000 simulations without control variates. However, if we value the cap at the exercise moment of the swaption, we get $\kappa \approx 200$, so we would only

[14] If valued at the exercise date, the se of the Bermudan is equal to 0.77 (see table 12.1). If we value it at T_{11} the se is 1.04.

[15] The simulated value of the cap will be lower than the analytical value.

need 500 simulation to get the same accuracy! We will have a closer look at this enormous improvement later on in this chapter. Note that the difference between valuing at T_e and T_{e+1} is very large, while we only value the cap one timestep later.

And what about valuing the control variate before T_e? As can be seen from the first bars of the figure, this is also not optimal. Of course, the larger u is, the larger κ, because $\min(T_e, T_u)$ converges to T_e: for $u = 9$ only those paths where the CLE is exercised at T_{10} or T_{11} or not at all are different.

The last interesting question is how to apply this to CLE's whose value is not known at the exercise date T_e, like the callable snowball. In that case, we can only value the CLE at T_{11}. The optimal valuation date for the cap is no longer T_e, but T_{11}, the same as the valuation date of the CLE. We can test it by valuing the swaption at time T_{11}, using the simulated Libors up to T_{11} instead of the analytical swap value at T_e. The standard error of the value is now larger compared to exercised at time T_e (se $= 1.04$ instead of 0.77 from table 12.1). As we would expect from the analysis above, it turns out that valuing the cap at T_{11} is much better ($\kappa = 85$), compared to valuing at T_e ($\kappa = 2.3$).

The same analysis as done here can be done for every other control variate instead of the cap. The results will be similar. From now on, I will always value the control variate at the same moment as the CLE.

12.4 Single control variates

Now we will have a look at the effectiveness of the different control variates proposed in section 12.2. First we will use the underlying swap as a control variate. Further 3 different European swaptions will be used: one expiring at T_1, one at T_6 and one at T_{12}. Finally we will again take the cap (see also the previous section). For all CV's the strike is equal to the strike of the Bermudan. The results can be found in figure 12.2.

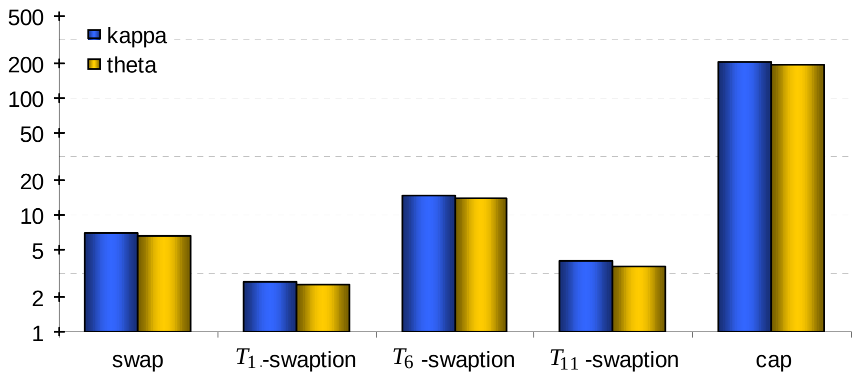


Figure 12.2: κ and θ for different control variates for the Bermudan payer swaption.

Note that the difference between θ and κ is very small. This means the extra time necessary to value the control variate for each path is very small compared to the total simulation time. The reason is that the lion's share of computation time is in the simulation of the paths. The pricing of the CV given this path is relatively fast, so this does hardly affect the performance. For swaptions the impact is relatively large, because the estimation of the swap rate volatility costs some time.

As can be expected, the swaption with maturity T_6, performs better than the one with maturity T_1 (to early) or T_{11} (to late, earlier cash flows do not occur). It turns out T_6 is in this case better than all other maturities T_i, but these are not displayed here. As mentioned before, because swaptions are not priced exactly in the LMM, we have to see whether this leads to inaccuracies. The estimated value of the Bermudan swaption is not significantly different when the swaptions are used as control variate, so it seems we can use the swaption as control variates without any problems

Even though the swap and European swaptions show significant improvements (up to a factor 15 for the T_6-swaption), they are by far outperformed by the cap. The correlation between the simulated values of the cap and Bermudan are 0.998, almost perfectly correlated. In the next section we will try to give an intuitive explanation of why the cap is such a good control variate for the Bermudan swaption.

12.5 Why does the cap perform so well?

Take a cap with strike R which is close to or (in this case) equal to the strike R_S of the Bermudan. First consider the case when Libor rates go up and will stay higher than R_S. In that case the cap pays $L_i - R$ and the Bermudan will be exercised and pays $L_i - R_S$. R and R_S are just constants, so these payments are perfectly correlated. Alternatively, suppose the Libor rates move, and stay, below R_S. Then the payoffs of the cap are zero and the Bermudan will not be exercised and will also have no payoffs. Again they are perfectly correlated. So in both extremes (which are also responsible for the largest part of the variance of the products value), they are perfectly correlated. Now suppose the Libor rates will fluctuate a little around R_S. In that case the cap will have a few small payoffs. The Bermudan will not be exercised, or, when exercised, will have positive as well as negative payments. Here we have a difference between the payoffs, but because the Libor rates are not fluctuating strongly, the impact will be relatively small. We can even go further: when the Libor rates first go down and afterwards go up, the payoffs of the cap and Bermudan will again be strongly correlated (the Bermudan will not be exercised until the rates go up, but also the cap does not have payoffs before exercise). The only large differences between the cap and Bermudan payoffs arise when the Libor rates go up first and afterwards go down or when they fluctuate strongly around R_S. In both cases the cap will have larger payoffs than the Bermudan.

Of course, this explanation is only meant to give an intuitive idea. We will come back to it in the next chapter.

12.6 Vector of control variates

It is also possible to take a vector of control variates instead of a single one, leading to a linear combination of products as control variate. The weights can be optimized to get a maximum correlation (see section 6.2.1). We will consider the following combinations of products: the three swaptions from the previous section (with expiry dates T_1, T_6, T_{11}), all swaptions ($T_1, .., T_{11}$), cap+swap, cap+T_6-swaption, cap+all swaptions. Finally we will decompose the cap into caplets and use them as a vector of control variates. Here, if all elements of the vector β would be equal to 1, we get the original cap. By allowing β to de different from 1, we will get an even higher correlation.

The results are in figure 12.3. We can see that using more swaptions increases the convergence speed. However, especially when all swaptions are used, the computation time also increases substantially. Using all swaptions instead of just the T_6-swaption

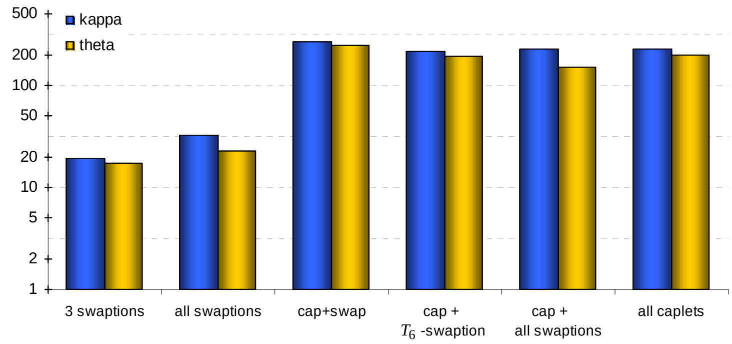


Figure 12.3: κ and θ for different linear combinations of control variates for the Bermudan payer swaption.

leads to a doubling of κ. Still, it is not as good as the single cap. Furthermore we see that adding other control variates to the cap does lead to some improvement, but it is not dramatically better. The swap seems to be the best choice, with a 30% improvement, compared to the single cap. For the combination of caps with European swaptions, the gain in convergence is not large enough to compensate for the extra computation time. Also decomposing the cap into caplets does not give strong improvements. The element β_i are all relatively close to 1, so the control variate is still almost equal to the cap.

Note that all control variates are strongly correlated. This induces the risk of multi-collinearity. However, for the examples above, this does not seem to be the case, because the vector of weights β does not show extreme values and the weights seem to be stable among different simulations.

12.7 Antithetic variates

Antithetic sampling (AS) is very easy to implement and widely used in practice. Moreover, it can easily be combined with the use of control variates. We would expect that the combination of the methods leads to an increased convergence. We use only half the amount of simulations (so 50000) as for the other control variates (because each simulation m already contains two valuations, for z_m and $-z_m$).

Figure 12.4 shows the results, together with the original results from figure 12.2, to make it easier to compare them. Only κ is given, because θ is almost the same. The first column shows the improvement in variance when no control variates are used: $\kappa = 2.6$. If the improvement of the use of antithetic sampling and control variates would be independent, the κ's of the control variates are multiplied by 2.6. Unfortunately, this does not seem to be the case For the cap, the improvement is still substantial, increasing κ from 200 to 300, so 50% faster. For the swaptions there is also some improvement, but small. For the swap however, including antithetic variables decreases the performance of the CV. The reason is probably that the difference between the swap and the Bermudan swaption payoffs is not monotonic (see also section 6.1). When β is smaller than 1, the difference between the swaption and swap is more or less symmetric around the fixed rate. This is a very important observation, because it shows we cannot use antithetic sampling for any situation and have to be careful with it.

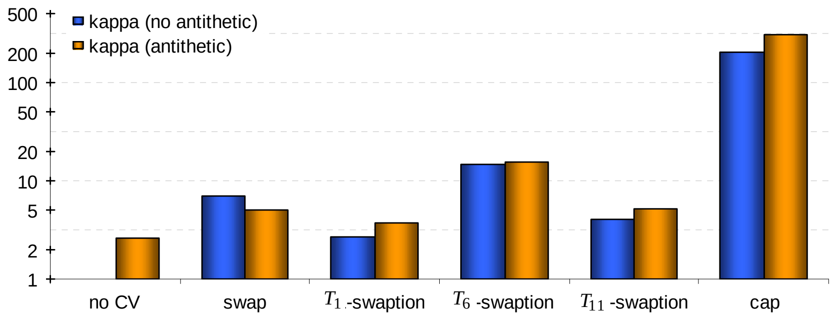
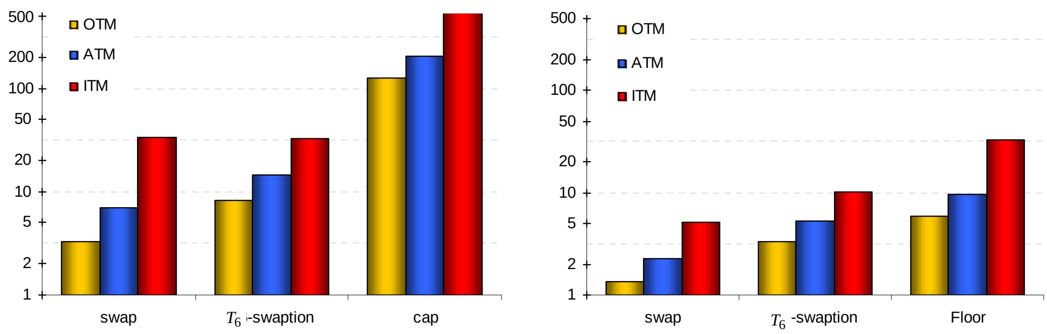


Figure 12.4: κ for different control variates for the Bermudan payer swaption, with and without the use of antithetic sampling.

12.8 Other swaptions

In the previous section we have produced results for the ATM Bermudan payer swaption. In this section we investigate whether the results are the same when we have a receiver swaption and/or the swaption is in- or out-of-the-money. Finally we pay attention to the case when the swaption is replaced by a cancellable swap.

First we have a look at the moneyness of the swaption. We will not repeat all results, but only look at the cap, swap and T_6-swaption. Because the differences between κ and θ are again very small, only the results for κ are given. See figure 12.5.



Bermudan payer swaption

Bermudan receiver swaption

Figure 12.5: κ and θ for different levels of moneyness

We see the control variates work better when the product is more in-the-money. When the products is in the money, it will be exercised early, so the product behaves more like a swap. Therefore the swap works better as a control variate. Also the European swaption looks more like a swap, which also explains why the effectiveness of the swap and T_6-swaption are almost equal for the ITM-swaption. The most important conclusion is that the relative differences are small: the cap still is by far the best control variate.

Now have a look at the receiver swaption. Note that the cap has been replaced by a floor and the payer swaption has been replaced by a receiver swaption. For the swap, there is no difference, because the payer swap is just -1 times the receiver swap (so also

β is multiplied by -1).

As we can see, the performance of the control variates is clearly worse for the receiver swaption. Instead of a variance reduction of order 100, the floor only gives a variance reduction of order 10. The reason is probably as follows. In the model the yield curve is upward sloping. This means that in general $L_i > L_j$ for $i > j$. This has a positive effect on the value payer swaption: When a payoff is positive, and we exercise, we can expect that future payoffs will generally also be positive. For the receiver the opposite holds. When we exercise, the value of later payoffs is expected to be lower than the current payoff. This gives rise to a larger probability of negative cash flows. Because the cap does not have any negative payoffs, this decreases the correlation with the cap as control variate. Another consequence is that the product will generally be exercised later. This reduces the correlation with the cap and swap. Besides the large decrease in effectiveness, the relative performance between control variates is more or less the same. The floor is still the best control variate.

Finally, what happens if we replace the swaption (callable swap) by a cancellable swap? From the relation between callables and cancellables (equation 8.2), we can deduce which choice of control variates we should take. The cancellable receiver is equal to the sum of the callable payer and the underlying receiver swap. For the first we can use the cap as control variate, as we have seen before. The swap already is analytical, so we can replicate it perfectly by itself. Therefore the swap+cap should be an effective control variate for the cancellable receiver. Note that the sum of a receiver swap and cap is equal to the value of a floor. So for the cancellable receiver swaption we expect that the use of a floor has about the same effect as using a cap for the callable payer swap. By the same argument we use a cap for a cancellable payer swap, which will be less effective, because the control variates for the callable receiver swaption are also inferior to those for the payer. Results, which will not be given here, confirm that the reasoning above is correct and give approximately the same results. Because the results will be the same, we will only focus on the callable swaps.

12.9 Preliminary conclusions

Several conclusions for the swaption can be drawn from the results in this chapter, which can also be applied to other products, as will be done in Chapter 14. First, it is optimal to value the CV at the same moment as the CLE. Second, the use of antithetic samples can improve the MC convergence. However, when used in combination with control variates, this is no longer guaranteed and the combined effect is not as large as the individual improvements. Therefore it is always important to be cautious and not use antithetics recklessly. Both conclusions can be expected to hold for other products as well.

With respect to control variates, we have seen that the cap is a very good control variate for the payer swaption. However, when we look at receiver swaptions, the improvement of using the floor as CV is much smaller, which can be explained by the upward sloping yield curve. Since the cap/floor always has the best performance, we will focus on it in the next chapter. We will try to obtain even better results by looking at other control variates, which are closely related to the cap.

The effectiveness of the cap is partly explained by the close resemblance with the swaption, so it will probably be less effective for other products. However for other CLE's, we can try to take the capped payoff of the underlying Libor exotic as control variate. This will be investigated in Chapter 14.

With respect to the results for the control variates, the conclusions are similar to those by Jensen and Svenstrup (2005). They looked at the Bermudan payer swaption in a model where the initial yield curve is flat. They found that the cap was a very good control variate. Since they looked at a flat term structure, we would expect the results to be somewhere between the results for the upward sloping and the downward sloping term structure (i.e. the receiver swaption), which is true indeed.

13 Improving the cap

In the previous chapter we saw that the cap[16] is a very good control variate for the Bermudan swaption. We only looked at a specific cap: the cap rate was equal to the swap rate and equal for all caplet-maturities. In this chapter we have a closer look at this and search for improvements. Especially we would like to improve the control variates for the receiver swaption, because the performance here is relatively poor. First we will have a closer look at the resemblance of the cap and swaption. Based on these observations, we will develop methods to improve the performance and finally we will test these on the Bermudan swaption.

13.1 Investigating the cap

We would like to optimize the correlation between the value of the Bermudan swaption and the cap. We can rewrite the former (equation 9.3) as:

$$B(T_0)\mathbb{E}^0 \sum_{i=1}^{N} \frac{(L_i - R_S)\,\mathbf{1}_{\{\tau<T_i\}}}{B(T_i)}, \tag{13.1}$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function and τ is the optimal stopping time.[17] The value of the cap is given by

$$B(T_0)\mathbb{E}^0 \sum_{i=1}^{N} \frac{[L_i - R]^+}{B(T_i)}. \tag{13.2}$$

If $R = R_S$ (as was the case in the previous chapter), for all i the cash flows are exactly the same, but the swaption only pays cash flows at dates $T_i > \tau$ and the cap only pays at dates T_i where $L_i > R$. Now we will focus on the individual cash flows at each date.

Let's have a closer look at the payoffs of the cap and the payer swaption. At time T_i, the cap pays

$$[L_i - R]^+ \equiv (L_i - R)\,\mathbf{1}_{\{L_i>R\}}.$$

The payoff of the swaption depends on whether it has been exercised. We can write it as

$$(L_i - R_S)\,\mathbf{1}_{\{\tau<T_i\}}.$$

From the two equations, we see the cap is a very good control variate when there holds $\mathbf{1}_{\{L_i>R\}} \approx \mathbf{1}_{\{\tau<T_i\}}$ for all i. Most of the time, the equality holds: if the cap is out of the money ($L_i < R$) it is never optimal to exercise. Moreover, we will exercise when the current and future caplets are in the money. There are two situations when the two indicators at T_i are not equal:

- The caplet is out-of-the-money ($L_i < R$), while the swaption has been exercised ($\tau < T_i$). This will happen when previous caplets where (deep) ITM, but current Libors are lower (generally when the term structure is downward sloping), so caplets with a later maturity are OTM.

[16] We will not keep saying cap for the payer swaption and floor for the receiver swaption. In general, everything that holds for one, also holds for the other. Usually I will only discuss the cap, but keep in mind that the same holds for the floor applied to the receiver swaption.

[17] Previously τ was the index of the exercise date. We will use both notations, since it is always clear which one is meant.

- The caplet is in the money ($L_i > R$) but the swaption has not yet been exercised ($\tau \geq T_i$). This will be the case when it is not optimal to exercise, even if the caplet is in the money, because forward Libors are not (so $L_j(T_i) < R$ for some $j > i$).

When the (initial) term structure is upward sloping, both possibilities will not occur very frequently. When we exercise (and thus $L_i > R$), also future caplets are generally in the money (so the first situation will not occur). Therefore it is generally optimal to exercise as soon as the next caplet is in the money (which excludes the second situation). This explains why the correlation between the cap and Bermudan payer swaption is very high.

Contrarily, when the term structure is inverted, these situations will occur more often. If we decide to exercise (thus the current caplet is ITM) there can still be future caplets that are OTM, because future Libor rates are expected to be lower (first situation). If these negative cash flows are relatively large, it will not be optimal to exercise as soon as the caplet becomes in the money, but only when it is deeply in the money ($L_i >> R$, second situation).

For the receiver swaption, we have to look at $R - L_i$ instead of $L_i - R$. Similar arguments hold. However, when the term structure if upward sloping, the payer swap payoff $L_i - R$ is increasing in i. However the receiver payoff $R - L_i$ is decreasing. Therefore an upward sloping yield curve for a receiver swaption has the same effect as a downward sloping yield curve for the payer swaption. As argued above, this will give a lower correlation. This explains the results from the previous chapter. So if our initial term structure would have been inverted, we would get very good results for the receiver swaption, but worse results for the payer swaption. As stated in the objectives (section 11.1), this is not what we want (even though a variance reduction factor of around 10 is still good). We would like to have a method that works for receivers as well payers, or equivalently, for normal and inverted term structures. Therefore we will try to improve the results in this chapter.

Now that we have identified the (dis)similarities between the cap and swaption, we can look for improvements. In both situations when $\mathbf{1}_{\{L_i > R\}} \neq \mathbf{1}_{\{\tau < T_i\}}$ we see that the caplet payoff is higher than the swaption payoff. The simplest way to compensate for this, is to increase the strike of the cap. We will investigate this in the next section.

13.2 Changing the strike

We investigate what happens to the convergence if we change the cap rate R of the cap/floor. Figure 13.1 shows κ as a function of the cap rate R.. The original results (where the cap is ATM) is marked by a dot.

First look at the payer swaption. We can see that we can increase κ to more than 500 if we increase the strike to 0.0345. For the receiver swaption, we see that the floor rate we used in the previous section ($R = 0.0322$) is much further away from optimal. If we change the cap rate to 0.026, we almost double κ (from 9 to 17). The explanation can be found in the arguments from the previous section. Because the yield curve is upward sloping, the future cash flows will generally be lower than the current. Therefore it will generally not be optimal to exercise as soon as $L_i < R$, but only when $L_i < R - \delta_i$ for some $\delta_i > 0$, which depends on the current term structure. Moreover, δ_i is increasing in i, because the probability that the swaptions has been exercised before T_i is an increasing function. Therefore we could expect to get an even higher correlation if we allow the floorlets to have different strikes.

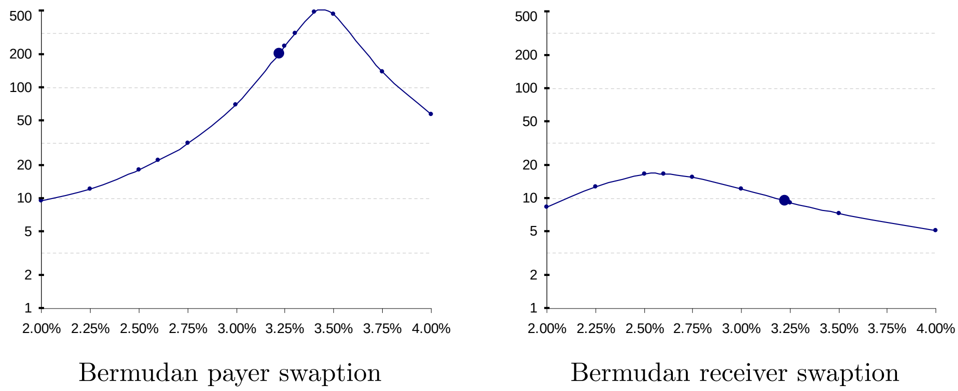


Figure 13.1: κ for the cap/floor as CV for the Bermudan payer swaption as function of the cap rate R.

As we could expect from the previous section, the optimal strike for the floor is further away from the Bermudan's fixed rate than for the cap. Of course, the optimal strike is larger than the fixed for the payer and smaller for the receiver swaption. Finally, we see that, especially for the payer swaption, κ deteriorates very fast when the strike is further away from the optimal strike.

The question, of course, is how the optimal strikes could be estimated. First we have a closer look at the cap/floor.

13.3 Decomposing the cap

Suppose we have a cap with strike R. At each cash flow date T_i it pays

$$(L_i - R)\,\mathbf{1}_{\{L_i > R\}}.$$

We can take a more general version of the cap, by letting R be different for each cash flow date ($R = R_i$), so the cap consists of caplets with a different strike.

$$(L_i - R_i)\,\mathbf{1}_{\{L_i > R_i\}}.$$

This caplet pays $L_i - R_i$ if $L_i > R_i$. We could equally well generalize this to a product that pays $L_i - R_i$ if $L_i > B_i$

$$(L_i - R_i)\,\mathbf{1}_{\{L_i > B_i\}}. \tag{13.3}$$

We will call this product a shifted cap (SCap). Moreover, we will call B_i the barrier to distinguish it from the strike R_i. We can decompose it in the following way:

$$\begin{aligned}(L_i - R_i)\,\mathbf{1}_{\{L_i > B_i\}} &= (L_i - B_i + B_i - R_i)\,\mathbf{1}_{\{L_i > B_i\}} \\ &= (L_i - B_i)\,\mathbf{1}_{\{L_i > B_i\}} + (B_i - R_i)\,\mathbf{1}_{\{L_i > B_i\}}. \end{aligned} \tag{13.4}$$

The first term is just a cap with strike B_i and the second term is $(B_i - R_i)$ times a digital with strike B_i. From this it follows that the value of this product can be computed from equations (1.10) and (1.11).

We would like to choose B_i and R_i in such a way that we get the highest correlation with the Bermudan swaption. We will investigate two different methods to do this. The first is by taking $R_i = R_S$ and trying to find B_i such that $\mathbf{1}_{\{\tau<T_i\}} \approx \mathbf{1}_{\{L_i>B_i\}}$. The second method tries to approximate the shifted cap with unknown optimal R_i and B_i by a linear combination of caplets and digitals with pre-determined strikes R_i. These methods will be described in the following sections.

13.4 Method 1: Shifted cap (regression)

The first method, which we will call the shifted cap or regression method, looks at each cash flow date separately. The cash flow of the Bermudan swaption and shifted cap at T_i are given by (see also equations 13.1 and 13.2):

$$\begin{aligned} CF_i^{BS} &= (L_i - R_S)\,\mathbf{1}_{\{\tau<T_i\}}, \\ CF_i^{CV} &= (L_i - R_i)\,\mathbf{1}_{\{L_i>B_i\}}. \end{aligned}$$

We try to approximate the cash flows of the swaption by setting $R_i = R_S$ for all i. What remains to be done is the determination of B_i. From the cash flows we can expect that the highest correlation will be found when B_i is chosen such that

$$\mathbf{1}_{\{L_i>B_i\}} \approx 1_{\{\tau\leq T_{i-1}\}}. \tag{13.5}$$

Since τ does not only depend on L_i, there will not exist a B_i giving an excellent fit. B_i can be determined by minimizing the (squared) difference between the two indicator functions over a set of paths. Note that for this method to be applicable to other products, it is important that the payoff at each time T_i $(L_i - R)$ is a monotonic function of the corresponding Libor rate L_i.

The following method will be used to determine B_i. The algorithm has some similarities with the Longstaff-Schwartz algorithm (section 9.3.1). We will look at the exercise strategy for the cancellable. Recall from section 9.3.1 that the product will be exercised at the first T_i where the continuation value is smaller than zero:

$$H_{i+1}(T_i) < 0.$$

From this we can define whether the product has been exercised before time T_i:

$$\tau \leq T_{i-1} \iff \min_{1\leq j\leq i-1} H_{j+1}(T_j) < 0.$$

Now write $\min\limits_i H \equiv \min_{1\leq j\leq i-1} H_{j+1}(T_j)$. Suppose we can estimate this by[18]

$$\min_i H \approx a_i - b_i L_i(T_{i-1}) \qquad (i \geq j+2). \tag{13.6}$$

Then the barrier for the caplet with cash flow date T_i is given by solving $\min_i H < 0$. The solution is given by:

$$L_i > \frac{a_i}{b_i}.$$

so that we have found the barrier: $B_i = \frac{a_i}{b_i}$. The only thing that remains to be done is to estimate a and b in equation (13.6). Just as in the Longstaff-Schwartz method, we do this by a cross-sectional regression. Because H_i is already known from the Longstaff-Schwartz optimization for the Bermudan swaption, we can simply compute $\min_i H$ for each path and use a linear regression to find a and b.

[18] The reason that there is a minus sign in front of β_i is that H is negatively correlated with L_i, so the minus sign ensures $\beta_i > 0$.

To summarize, we have tried to estimate (13.5) in the following way

$$1_{\{\tau \leq T_{i-1}\}} = 1_{\{\min_i H<0\}} \approx 1_{\{a-b_i L_i<0\}} = 1_{\{L_i>a_i/b_i\}} = 1_{\{L_i>B_i\}}.$$

The method can be implemented fairly easy. A weak point is that we need an extra regression to estimate B_i. Moreover, it can be questioned whether the approximation from equation (13.6) is by any means realistic. We can think of several ways to try to improve things. See appendix D for a few possibilities. Because these alternatives did not lead to better results, they will not be discussed any further.

Note that θ for the regression method can not easily be compared to the other ones. This is because the time necessary to estimate the strikes is relatively larger when we use fewer simulations (because it is constant). If we would use less simulation, θ will also decrease.

13.5 Method 2: vector of caps

Another way to estimate the optimal cap is to estimate it by a linear combination of other caps. As explained in section 6.2.1, it is easy to use a linear combination of products as control variate. Moreover, the simulation determines the optimal weight factors. Because we do not know the optimal strike, but we do know it will be around the Bermudan strike R_S, we can use a vector of caps with strikes around R_S as an approximation. The following two examples illustrate the idea.

Suppose the optimal control variate is a shifted cap, which is defined by its payoffs at T_i from equation (13.3):

$$(L_i - R_i)\,\mathbf{1}_{\{L_i>B_i\}}.$$

where R_i and B_i are the optimal parameters, resulting in the highest correlation. Furthermore suppose we would know the optimal value for B_i but we do not know the optimal value for R_i. We can estimate it by using equation (13.4):

$$\begin{aligned}(L_i - R_i)\,\mathbf{1}_{\{L_i>B_i\}} &= (L_i - B_i)\,\mathbf{1}_{\{L_i>B_i\}} + (B_i - R_i)\,\mathbf{1}_{\{L_i>B_i\}} \\ &= \mathrm{cap}_{B_i} + (B_i - R_i)\,\mathrm{digicap}_{B_i}\,. \end{aligned} \tag{13.7}$$

Now we use a vector of two control variates consisting of this cap and digital. If we then run a simulation, we can optimize the weights $\beta = [\beta_1, \beta_2]$ as described in section 6.2.2. This gives us the following linear combination as control variate:

$$\begin{aligned} CV &= \beta_1\,\mathrm{cap}_{B_i} + \beta_2\,\mathrm{digicap}_{B_i} \\ &= \beta_1 \left(\mathrm{cap}_{B_i} + \frac{\beta_2}{\beta_1}\,\mathrm{digicap}_{B_i}\right). \end{aligned}$$

Setting the term between brackets equal to equation (13.7), we can compute R_i:

$$R_i = B_i - \frac{\beta_2}{\beta_1}.$$

Because we can estimate β in the same simulation, it does not require extra time to estimate R_i. So we have estimated the optimal control variate by a linear combination two other control variates.

Of course, we do not know the optimal value for B_i so we cannot use this method directly. We will use the same idea for the barrier B_i. Unfortunately, it is not possible to give the payoff as linear combination of caps independent of B_i as we did before for R_i. However, we could for example try to use the following approximation:

$$\beta \left[L_i - R^*\right]^+ \approx a \left[L_i - R^1\right]^+ + b \left[L_i - R^2\right]^+ . \tag{13.8}$$

The left hand side is the 'optimal cap' with unknown strike R^*, which we try to approximate by the two caps on the right hand side. We choose R^1 and R^2 beforehand, such that $R_1 < R_2$. We can write out the payoff of the caplets on the right hand side:

$$a \left[L_i - R^1\right]^+ + b \left[L_i - R^2\right]^+ = \begin{cases} 0 & L_i \leq R^1 \\ a\left(L_i - R^1\right) & R^1 < L_i \leq R^2 \\ (a+b)\, L_i - \left(aR^1 + bR^2\right) & R^2 < L_i \end{cases} .$$

We can use this to approximate the value of a single caplet with an unknown strike R^*, whose payoff is given by:

$$\beta \left[L_i - R^*\right]^+ = \begin{cases} 0 & L_i \leq R^* \\ \beta\left(L_i - R^*\right) & R^* < L_i \end{cases} .$$

Assume $R_1 < R^* < R_2$. Then there hold that equation (13.8) holds for $L_i \leq R^1$. It will also hold for $R^2 < L_i$ if a and b solve:

$$\begin{aligned} a + b &= \beta, \\ aR^1 + bR^2 &= \beta R^*. \end{aligned} \tag{13.9}$$

See figure (13.2). Now we can use the two caplets with strike R^1 and R^2 as control

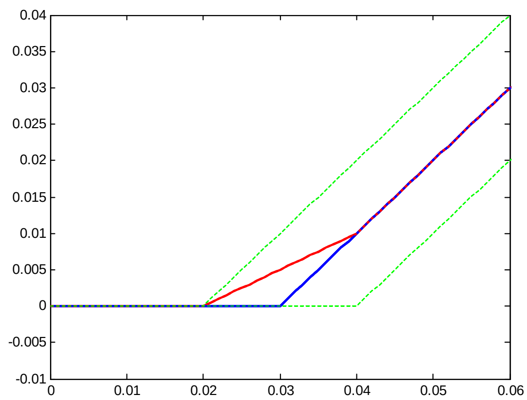


Figure 13.2: Approximation of caplet with strike 3% (blue) by two caplets with strikes 2% and 4% (green). The linear combination of caplets that solves equations (13.9) is given by the red line.

variates. The variance minimization will give a and b. Then we can estimate the 'optimal control variate' from:

$$\begin{aligned} \beta &= a + b, \\ R^* &= \frac{aR^1 + bR^2}{a+b}. \end{aligned}$$

The caplet with this strike (R^*) has the same payoff as the two caplets for $L_i < R^1$ and $R^2 < L_i$. Again we have estimated an unknown caplet by two other caplets. We do have to choose R^1 and R^2 in advance. An intuitive choice is $R^1 = R_S - 0.01$ and $R^2 = R_S + 0.01$, where R_S is the fixed rate of the swaption. Because we would expect the optimal strike is close to the Bermudan's fixed rate, we can expect that these are suitable lower and upper bounds.

These two examples are meant to illustrate the general idea, that we can use a set of caps and digitals with different strikes as control variate. An advantage is that it is not necessary to estimate anything beforehand; the optimal weight vector β is estimated by the simulation. We do not have to restrict to a set of two caplets and/or digitals, we can use any number of caplets, with different strikes we want. However, there is a risk of multi-collinearity, because caplets with almost equal strike are strongly correlated. In the following we choose to use three caplets, with strike $R_S - 0.01, R_S$ and $R_S + 0.01$.

These examples only looked at single caplet payoffs, but we have to find control variates that work for all cash flow dates (i.e. the whole swaption). As we noticed before, the optimal strike/barrier for each caplet will not be the same, but will be a decreasing (caplet) or increasing (floorlet) function of the maturity date T_i. If we just take a linear combination of caps however, it is not possible to differentiate the strikes between different cash flow dates. To make it possible to get different strikes for different maturities, we have to take individual caplets as control variates. Suppose we would use caplets with three strikes to approximate the 'true' caplet. When there are 11 cash flow dates, we would get $11 \times 3 = 33$ control variates. This may lead to instability in the determination of β. If we have F control variates we have to estimate $F(F+1)/2$ elements in the variance-covariance matrix and F covariances with the swaption, where we even ignore the use of digitals. We could also choose to steer the middle course, by decomposing the cap into a set of caps, each with different cash flow dates. We could for example use three caps for each strike: one with cash flow dates $\{T_1, T_2, T_3\}$, one with $\{T_4, T_5, T_6, T_7\}$ and the last one with $\{T_8, T_9, T_{10}, T_{11}\}$. In this way we get $3 \times 3 = 9$ caps as control variate.

13.6 Results

In this section we will compare the results for the two methods. We will compare the following control variates:

- cap: the base case from the previous section
- shifted: as in section 13.4
- 3 caps: a linear combination of three caps, with strikes R_S, $R_S + 1\%$, $R_S - 1\%$.
- 3 caps + 3 digitals: the same caps above plus three digitals with the same strikes
- 3x3 caps: We divide the cap in three sub-caps: The first consists of only caplets with maturity T_1, T_2, T_3, the second at maturities T_4, T_5, T_6, T_7 and the last with maturities T_8, T_9, T_{10}, T_{11}. We do this for each of the 3 caps with strikes R_S, $R_S + 1\%$, $R_S - 1\%$., so we get in total 9 caps.
- all caplets: For each of the three strikes we take all caplets separately, so we get 33 control variates.

The results can be found in figure 13.3. For the shifted cap strategy we have used 1000 paths to estimate the strikes.

As we can see, all proposed improvements outperform the original ATM cap. The improvements for the receiver swaption are larger than for the payer swaption. The

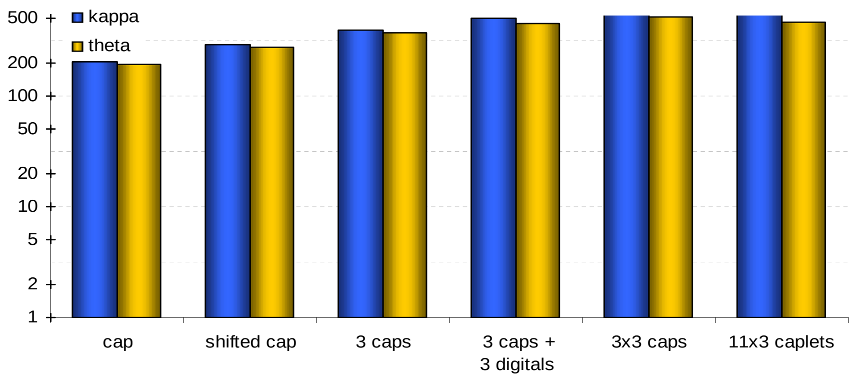


Figure 13.3: Bermudan payer swaption: κ and θ for different alternatives for the cap.

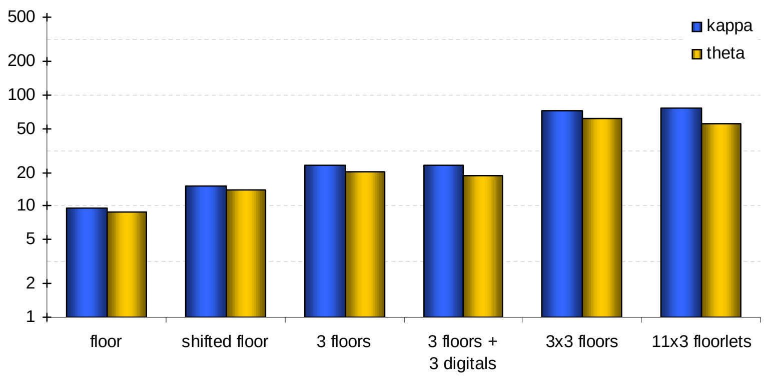


Figure 13.4: Bermudan receiver swaption: κ and θ for different alternatives for the cap.

regression method is somewhat disappointing: it looks like estimating the strike by this regression does not yield a good approximation, since there is hardly any improvement. For the second method we see that adding three digitals is of hardly any use. This also turns out to be the case for combinations of digitals with other caps, so we will not investigate the use of digitals any further. The decomposition of the floor into subfloors does lead to a significant improvement. Furthermore we see that the decomposition in 3 subfloors is almost as good as decomposing the floor in all floorlets. Because the former is more stable, we prefer this one. Moreover, because we have to compute a larger variance-covariance matrix, the method with all floorets separately is also slower (even though we only compute the same number of caplet values). Please recall from the previous chapter that decomposing a single cap into caplets did not lead to an significant increase in performance.

For the 3x3 floors we find that κ is more than 70. Still not as good as for the payer swaption, but over 7 times faster than the single floor!

The results for the OTM and ITM swaptions are very similar to the results presented for the ATM swaption. Therefore we will not present these here.

14 Other CLE's

In Chapter 12 we found the cap to be a very efficient control variate for the payer swaption. In the previous chapter we showed that we can get even better results by looking at other 'cap-like' products. With these improvements, also the results for receiver swaptions are very good. It would be very advantageous if we could apply a similar procedure to a wide range of callable Libor exotics. This is the subject of this chapter. First we look at the callable inverse floater. Even though the coupon is more complicated, the underlying Libor exotic can still be priced analytically. Furthermore, to test a product whose underlying cannot be priced analytically, we will investigate the cancellable snowball. Moreover, the snowball's underlying is path-dependent. If we are able to find efficient control variates for these products, we can expect we can also extend it to other products.

The main idea is as follows. The payer swaptions payoff, when exercised, is equal to the payoff of the underlying swap $L_i - R$. The cap is just the capped payoff of this swap $[L_i - R]^+$. We will extend this idea to other product by taking the capped payoff of the underlying Libor exotic as control variate. This can only be used when this capped payoff can be valued analytically, so we need some approximation when this is not possible.

For both CLE's, we start by describing how we can cap the payoff of the underlying. Next we describe which control variates we will investigate, similar to those from the previous chapter. Finally we give results where we compare the effectiveness of the control variates.

14.1 Callable Inverse Floater

We will first look at the payer CIF described in section 11.3.2.

14.1.1 Capping the payoff

The payoff of the underlying Libor exotic (the inverse floater) at time T_i is given by equation (11.2), ignoring the daycount fraction α_i:

$$L_i - f - [R - f - L_i]^+ . \tag{14.1}$$

As indicated above, we 'cap' the payoff to get the following control variate:

$$\left(L_i - f - [R - f - L_i]^+\right)^+ = \begin{cases} 0 & L_i < R - f, L < \frac{R}{2} \\ 2L - R & L_i < R - f, L > \frac{R}{2} \\ 0 & L_i > R - f, L < f \\ L_i - f & L_i > R - f, L > f \end{cases} . \tag{14.2}$$

Now we use the fact that in this case $f \leq R/2$[19]. This can be rewritten as (see figure 14.1):

$$2\left[L_i - \frac{R}{2}\right]^+ - [L_i - R + f_i]^+ , \tag{14.3}$$

which is just the difference of two caplets. The use of this payoff as control variate corresponds to the use of a simple cap as control variate for the swaption.

[19] If $f > R/2$ we could rewrite the capped payoff as $[L_i - f]^+$, so again we could use a single cap as control variate. If the product characteristics would be different from section 11.3.2 ($c < \infty$) we would also get a different decomposition.

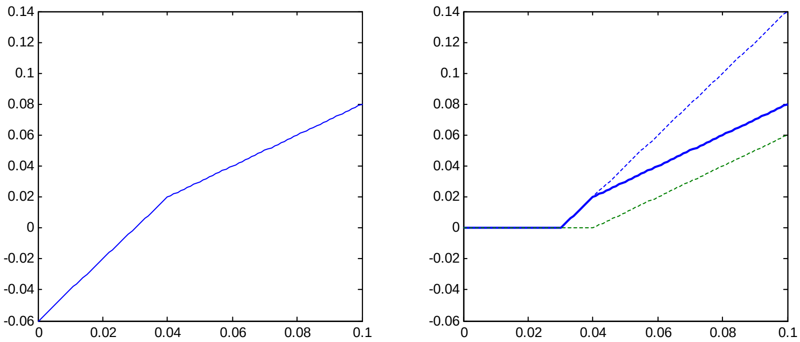


Figure 14.1: left: payoff of the inverse floater as a function of L_i (eq. 14.1). Right: capped payoff (eq. 14.2, solid) and the decomposition in caplets $2\left[L_i - \frac{R}{2}\right]^+$ (upper) and $[L_i - R + f_i]^+$ (lower). The capped payoff is the difference between these (eq. 14.3).

We can perform the same analysis for the receiver callable inverse floater. Now the capped cash flow is given by:

$$\left([R-f-L_i]^+ - L_i + f\right)^+ = \begin{cases} R-2L_i & L_i < R-f, L < \frac{R}{2} \\ 0 & L_i < R-f, L > \frac{R}{2} \\ L_i - f & L_i > R-f, L < f \\ 0 & L_i > R-f, L > f \end{cases} .$$

Because $f \leq R/2$, the third possibility ($L_i > R-f, L < f$) does not exist, so the cash flow is simple equal to a single floorlet:

$$2\left[\frac{R}{2} - L_i\right]^+ .$$

14.1.2 Control variates

We will investigate the variance reduction by the use of antithetic sampling as well as four different control variates. The first one is the underlying inverse floater (equation 14.1), which is the same as using the underlying swap as control variate for the Bermudan swaption. Next we look at the cap, defined as the capped payoff given in equation (14.3). Furthermore, we use two control variates suggested in the previous section. The first one is the shifted cap, where the strikes B_i are estimated by a regression (section 13.4). The other is the vector of 3x3 caps as proposed in section 13.5, where the caps are replaced by the capped payoff (equation 14.3).

For the shifted cap, we replace the capped payoff by:

$$\left(L_i - f - [R-f-L_i]^+\right) 1_{\{L_i > B_i\}}.$$

Just like the capped payoff, we can also rewrite this payoff. If $B_i > R-f$ we simply get $(L_i - f)\, 1_{\{L_i > B_i\}}$, which is just a shifted cap. For $B_i < R-f$ (and again $f \leq R/2$) it is

$$2\left(L_i - \frac{R}{2}\right) 1_{\{L_i > B_i\}} - [L_i - R + f]^+ .$$

Together this leads to:

$$2\left(L_i - \frac{R}{2}\right) 1_{\{L_i > B_i\}} - (L_i - R + f)\, 1_{\{L_i > \max(B_i, R-f)\}}.$$

We can estimate B_i as described in section 13.4.

For the 3x3 caps we use the same cash flow dates as in the previous chapter (T_1, T_2, T_3 for the first, T_4, T_5, T_6, T_7 for the second, T_8, T_9, T_{10}, T_{11} for the last cap). For the three different strikes, we use the following payoffs:

$$\begin{aligned}
CV_1 &: 2\left[L_i - \frac{R}{2} + \delta\right]^+ - [L_i - R + f_i - \delta]^+,\\
CV_2 &: 2\left[L_i - \frac{R}{2}\right]^+ - [L_i - R + f_i]^+,\\
CV_3 &: 2\left[L_i - \frac{R}{2} - \delta\right]^+ - [L_i - R + f_i + \delta]^+.
\end{aligned}$$

where $\delta = 1\%$ The cash flows as a function of L_i are given in figure 14.2

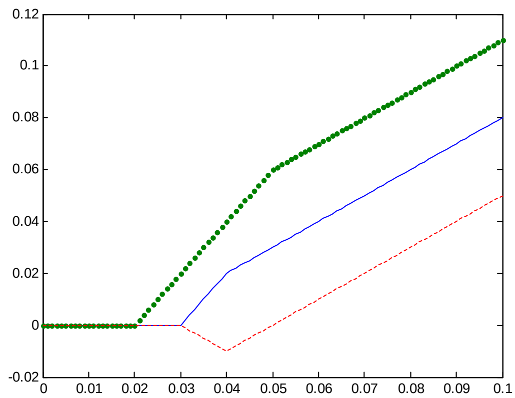


Figure 14.2: capped CIF payoff with different strikes

The estimation of the alternative CV's for the receiver are done in the same way (where of course caps are replaced by floors).

14.1.3 Results

The results are in figure 14.3 (payer) and 14.4 (receiver).

We see the results are similar to the Bermudan swaption. For both payer and receiver, antithetic sampling and the underlying as control variate only lead to minor improvements. For the payer callable inverse floater, the capped payoff works very good, reducing variance by a factor 100. Furthermore, the 3x3 caps do even more than two times better. For the receiver CIF the floor is again much better then the underlying or antithetic sampling, but the variance reduction is again only a factor 10. However we can improve this considerably by taking the 3x3 floors.

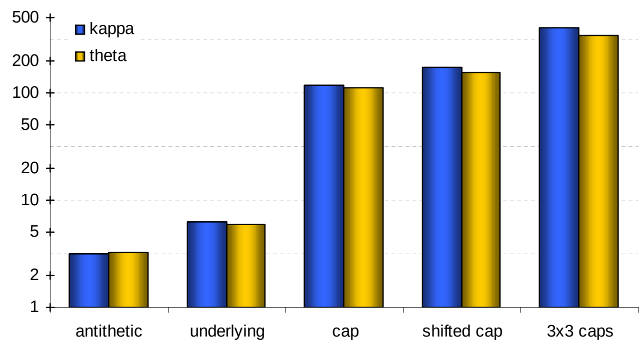


Figure 14.3: κ and θ for control variates for the ATM payer CIF

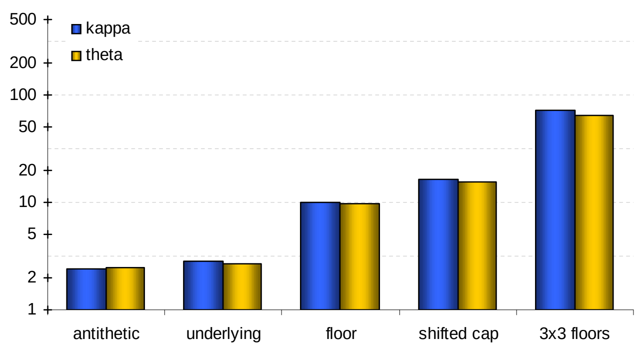


Figure 14.4: κ and θ for control variates for the ATM receiver CIF

Overall the results are very good. The results from the Bermudan can easily be extended to the CIF. We will not investigate the ITM and OTM CIF, but we can expect the results would be the same.

14.2 Snowball

Finally, we look at the cancellable snowball. Because the underlying is path-dependent, we cannot directly take the (capped) underlying as control variate. We will suggest methods to overcome this problem.

14.2.1 Capping the payoff

For the payer snowball, the cash flows of the underlying at T_i $(2 \leq i \leq N)$ are given by:

$$L_i - [C_{i-1} + A - L_i]^+ , \tag{14.4}$$

where C_i is computed from

$$\begin{aligned} C_i &= [C_{i-1} + A - L_i]^+ , \\ C_1 &= 0.0135. \end{aligned} \tag{14.5}$$

Suppose we would know the previous coupon C_{i-1}. Then the snowball cash flow is equal to the cash flow of an inverse floater, with $f = 0$ and $R = C_{i-1} + A$. For the CIF we know that we can use the capped cash flow as control variate. Capping this payoff gives

$$2\left[L_i - \frac{C_{i-1}+A}{2}\right]^+ - \left[L_i - (C_{i-1}+A)\right]^+ . \tag{14.6}$$

For the receiver this is replaced by:

$$2\left[\frac{C_{i-1}+A}{2} - L_i\right]^+ .$$

We cannot use this directly, because we do not know the previous coupon for the future cash flow dates. At T_i we know the cash flow at T_{i+1}, so we know the strike at T_{i+2}. All other strikes are unknown. A natural way to proceed is to estimate the previous strike C_{i-1} for each coupon by $\widetilde{C}_i$ at T_0. Then we can replace the control variate payoff (14.6) by:

$$2\left[L_i - \frac{\widetilde{C}_{i-1}+A}{2}\right]^+ - \left[L_i - \left(\widetilde{C}_{i-1}+A\right)\right]^+ , \tag{14.7}$$

and similar for the receiver. This can be used as a control variate. There are several ways to estimate the previous coupons $\widetilde{C}_i$:

- assume all strikes are equal to the first strike: $\widetilde{C}_i = C_1$ for all i. In a different context, this approximation is also used by Bender et al. (2005),
- assume the current forward rate would be realized, so $L_i(T_{i-1}) = L_i(T_0)$. Then the strikes can simply be computed recursively from equation (14.5):

$$\widetilde{C}_i = \left[\widetilde{C}_{i-1} + A - L_i(T_0)\right]^+ ,$$

- use a Monte Carlo simulation to estimate C_i.

We will use the second method, because it turns out to give the best results. Of course, the Monte Carlo estimate converges to the average strike, but this requires an extra simulation to estimate it. Moreover the average strike is not necessarily the best estimate for the strike, because there are large positive outliers. The strike is bounded below by 0, but when the Libors are low, it can become very large (the snowball turns into an avalanche).

14.2.2 Control variates

We will look at the same control variates as for the CIF. Because the underlying is no longer analytical, we will leave that one out.[20] What remains is antithetic sampling and three control variates: the cap (the capped payoff), the shifted cap and the 3x3 caps. We will also take into account two other control variates, that take into account the path-dependency of the snowball. We will discuss these now.

Whatever technique we use, at T_0 our estimate of the strike $\widetilde{C}_i$ in the future for a given path will be poor, because we cannot use any path information. At later date T_j we can make a better estimate of the strike for $i > j$, since we have more information

[20] We could of course take an approximation, in the same way as for the capped payoff, but because the underlying did not perform very good for the other CLE's, it will probably not be worth the effort.

(being the current coupon and forward Libors at $L_i(T_j)$). Therefore we will also look at a dynamic strategy, where we update the strikes at each timestep.

The control variate now becomes a self-financing portfolio, where we update the product at each timestep:

- At T_0 estimate all strikes $\widetilde{C}_i$ ($i > 1$), by assuming the current term structure is realized ($L_i(T_{i-1}) = L_i(T_0)$ for all i). Buy all the caplets from equation (14.6).
- At the next date T_j one of the caplets payoff (the one maturing at T_{j+1}) is known exactly. We estimate the other strikes C_i ($i > j+1$) again, by using the current coupon C_j and current term structure $L_i(T_j)$. We sell all the caplets from the previous step and buy the caplets with the new strike. The difference in value (plus the value of the caplet with maturity T_{j+1}) is the cash flow at this date (which is equal to lending or borrowing the money against the risk free rate to get a self-financing portfolio).
- Repeat the previous step until the snowball is exercised. At the exercise date, value all the remaining caplets.

This method enables us to take the path-dependency of the snowball into account. The drawback is that we have to value all the caplets at each date, so this will make the method slower. We call this the dynamic control variate. Moreover we will look at the corresponding dynamic 3x3 caps. We will not use the shifted cap (regression) strategy in combination with the dynamic CV, because then we would have to estimate the barrier at each date for each path (see also appendix D for a dynamic extension of the shifted cap).

14.2.3 Results

For the control variates described in the previous section, the results are given in figure 14.5 (cancellable receiver) and figure 14.6 (cancellable payer). As argued in section 12.8, the cancellable receiver can be expected to give similar results as the callable payer. Therefore we will first discuss the receiver.

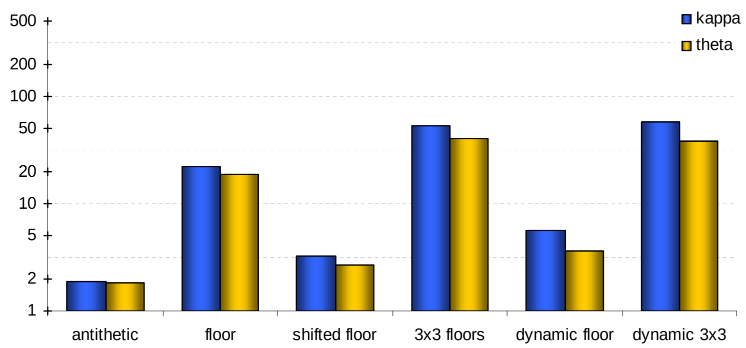


Figure 14.5: κ and θ for different control variates applied to the cancellable receiver snowball

We see some very interesting results. The floor is a good control variate ($\kappa = 20$). The use of a regression to estimate the optimal barriers performs very disappointing, while the 3x3 floors do give a significant improvement. The dynamic strategy is also very disappointing. However, if we use it in combination with 3x3 floors, the results are satisfactory.

It is hard to find an explanation why some methods perform very bad, while the original floor has a relatively good performance. We would expect that both the shifted cap as well as the dynamic floor would perform as least as good as the simple floor (even though a little slower). Probably the reason is the very asymmetrical payoff of the underlying of the snowball. The coupon payments are bounded from below by zero, but could become very large (the avalanche). Therefore it could be possible that the optimal exercise strategy is non-trivial. We can expect not to cancel product, even when payoffs are negative, because if we cancel we throw away a large upside potential. Somehow the regression technique seems not to be able to reproduce this effect. This does however not explain why the floor performs very good.

Now let's have a look at the cancellable payer. We see that the cap is worse compared to the floor for the receiver, just as we would expect from the results from the swaption and callable inverse floater. The shifted cap does not give any improvements (at least it is not worse than the cap this time). Once more, the 3x3 cap is clearly better. The dynamic strategy is also doing pretty good, but the improvements are partly undone by the extra computation time.

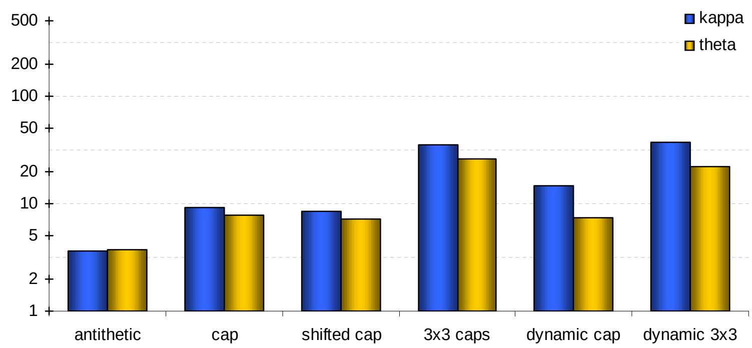


Figure 14.6: κ and θ for different control variates applied to the cancellable payer snowball

We can conclude that, even though the results are not as good as for the swaption and CIF, we have still found a large decrease in standard error by using the capped payoff as control variate. Taking the 3x3 caps gives even better results, while the shifted cap is very disappointing.

15 Conclusions

In Part IV of this thesis we have investigated the use of control variates to reduce the standard error of the Monte Carlo estimate of the price of callable Libor exotics.

We have investigated several control variates for the Bermudan swaption. We found that the cap (floor) is by far the best control variate for the Bermudan payer (receiver) swaption. Moreover we discovered that the shape of the term structure has a strong influence on the effectiveness of the cap, where the relative variance reduction ranges from a factor 10 to 200 for at-the-money Bermudan swaptions.

Next we have given an intuitive explanation why the cap is such a good control variate and investigated how to find even better control variates. We have proposed several methods to take other, similar, control variates that are more robust to the shape of the yield curve. Taking a vector of caps with different strikes and different cash flow dates turned out to lead to the best results, increasing the variance reductions to a factor 30 to 600 for at-the-money Bermudans.

Finally we extended the idea of the cap to other callable Libor exotics by taking the capped payoff of the underlying as control variate. For the callable inverse floater the results were very good, and very similar to the Bermudan swaption. The cancellable snowball is more complicated because the cash flows are path dependent. Still the use of a vector of caps showed a significant variance reduction (factor 30).

15.1 Suggestions for further research

There are numerous ways to extend the research presented in this thesis:

- Other improvements for the cap. The two methods suggested in Chapter 13 clearly lead to improvement compared to the plain-vanilla cap as control variate. Certainly, there will exist other ways to develop efficient cap-like control variates. Maybe it will be possible to improve on the methods we suggested.
- Other parameters. In this thesis we concentrated on a single set of parameters. We also tested the model on a different term and volatility structure, which yielded similar results. More research is necessary to see how the effectiveness of the control variates is affected by model parameters. Moreover, we only looked at CLE's maturing in six years ($T_N = 6$). For longer maturities, we would expect the cap to be less effective. Probably the improvements suggested in Chapter 13 will do relatively better, because the optimal strikes for different maturities will vary more for these products.
- Other CLE's. We have investigated three different CLE's. The results of Chapter 14 indicate that the method we presented can be extended to a wide range of CLE's. Further research has to show whether the method can be applied to other non-trivial CLE's like range-accruals or products with a TARN (Target Redemption Note) structure.
- Greeks. Besides obtaining the price of CLE's, Monte Carlo is also used to obtain Greeks, the partial derivatives of the price with respect to different uderlyings (see for example Glasserman and Zhao, 1999; Piterbarg, 2004). We did not investigate the computation of Greeks in this thesis, but it could be very interesting to see whether (and if so, how) control variates can be used to compute Greeks more efficiently.
- Application to other models. The method we have presented takes advantage of the similarities in payoff between callable Libor exotics and their capped underlyings. The

method is not restricted to the Libor Market Model. It might be interesting to see whether we can apply the same idea to other callable products and/or other models.

- Extension to stochastistic volatility. In this thesis we implemented the Libor Market model where volatilities are assumed to be deterministic functions of time. It is more realistic to assume that the volatilities themselves are also stochastic. Then we do not know the Black volatility exactly and have to compute caplet prices numerically. It would be interesting if we could somehow apply the control variates in these situation.
- Other variance reduction techniques. As mentioned in section 6.3, low-discrepancy sequences can be another effective way to reduce the standard error. Probably a combination of quasi-random numbers with control variates can give even better results.

Looking at this list, we could conclude that we are just getting started!

References

Andersen, L. (1999). A simple aproach to the pricing of bermudan swaption in the multi-factor libor market model, *Journal of Computational Finance* **3: 2**: 5–32.

Andersen, L. and Broadie, M. (2004). A primal-dual simulation algorithm for pricing multi-dimensional american options, *Management Science* **50: 9**: 1222–1234.

Baxter, M. and Rennie, A. (1996). *Financial Calculus, An introduction to derivative pricing*, Cambridge University Press.

Bender, C., Kolodko, A. and Schoenmakers, J. (2005). Iterating snowballs and related path dependent callables in a multi-factor libor model, *Weierstrass Institute Berlin* p. Preprint No. 1061.

Bender, C., Kolodko, A. and Schoenmakers, J. (2006). Iterating cancelable snowballs and related exotics in a many-factor libor model, *Risk* **to appear**.

Björk, T. (2004). *Arbitrage Theory in Continuous Time*, second edn, Oxford University Press.

Black, F. (1976). The pricing of commodity contracts, *Journal of Financial Economics* **3**: 167–179.

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

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 637–659.

Boyle, P. (1977). Options: a monte carlo approach, *Journal of Financial Economics* **4**: 323–338.

Boyle, P., Broadie, M. and Glassermann, P. (1997). Monte carlo methods for security pricing, *Journal of Economic Dynamics and Control* **21(8-9)**: 1267–1322.

Brace, A., Gatarek, D. and Musiela, M. (1997). The market model of interest rate dynamics, *Mathematical Finance* **7: 2**: 127–155.

Brace, A. and Womersley, R. (2000). Exact fit to the swaption volatility matrix using semidefinite programming, *Working paper presented at ICBI Global Derivatives Conference* .

Brigo, D. and Mercurio, F. (2001). *Interest Rate Models: Theory and Practice*, Springer Verlag.

Brzezniak, Z. and Zastawniak, T. (1998). *Basic stochastic processes: a course through exercses*, Springer Verlag.

Cox, J., Ingersoll, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica* **53**: 385–407.

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

Glasserman, P. (2003). *Monte Carlo methods in financial engineering*, Springer Verlag.

Glasserman, P. and Zhao, X. (1999). Fast greeks by simulation in forward libor models, *Journal of Computational Finance* **3**: 5–39.

Glasserman, P. and Zhao, X. (2000). Arbitrage-free discretization of lognormal forward libor and swap rate models, *Finance and Stochastic* **4**: 35–68.

Harrison, J. and Kreps, J. (1979). Martingales and aribtrage in multiperiod securities markets, *Journal of Economic Theory* **11**: 418–443.

Haugh, M. and Kogan, L. (2004). Pricing american options: A duality approach, *Operations Research* **52**: 258–270.

Heath, D., Jarrow, R. and Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology, *Econometrica* **60: 1**: 77–105.

Ho, T. S. Y. and Lee, S. (1986). Term structure movements and pricing interest rate contingent claims, *Journal of Finance* **41**: 1011–1029.

Hull, J. (2003). *Option, Futures and Other Derivatives*, fifth edn, Prentice Hall.

Hull, J. and White, A. (1990). Pricing interest rate derivative securities, *Review of Financial Studies* **3: 4**: 573–592.

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

Hull, J. and White, A. (1999). Forward rate volatilities, swap rate volatilities, and the implementation of the libor market model, *Journal of Fixed Income* **10: 3**: 46–62.

Hunter, C., Jäckel, P. and Joshi, M. (2001). Getting the drift, *Risk Magazine* **July**.

Jamshidian, F. (1997). Libor and swap market models and measures, *Finance and Stochastics* **1**: 293–330.

Jensen, M. and Svenstrup, M. (2005). Efficient control variates and strategies for bermudan swaptions in a libor market model, *Journal of Derivatives* **12:4**: 20–33.

Joshi, M. and Kainth, D. (2005). Enhancements to the longstaff-schwartz algorithm for bounding callable libor exotics, *Technical report*, QuARC, Royal Bank of Scotland.

Joshi, M. and Stacey, A. (2006). New and robust drift approximations for the libor market model, *Working Paper* .

Jäckel, P. (2002). *Monte Carlo Methods in Finance*, John Wiley & Sons Ltd.

Jäckel, P. and Rebonato, R. (2003). The link between caplet and swaption volatilities in a brace-gatarek-musiela/jamshidian framework: approximate solutions and empirical evidence, *Journal of Computational Finance* **6**: 35–45.

Kolodko, A. and Schoenmakers, J. (2006). Iterative construction of the optimal bermudan stopping time, *Finance and Stochastics* **10**: 27–49.

Longstaff, F. and Schwartz, E. (2001). Valuing american options by simulation: A simple least squares approach, *Review of Financial Studies* **14: 1**: 113–147.

Merton, R. (1973). Theory of rational option pricing, *Bell Journal of Economics and Management Science* **4**: 141–183.

Miltersen, K., Sandmann, K. and Sondermann, D. (1997). Closed form solutions for term structure derivatives with lognormal interest rates, *Journal of Finance* **52: 1**: 409–430.

Pietersz, R., Pelsser, A. and van Regenmortel, M. (2004). Fast drift-approximated pricing in the bgm model, *Journal of Computational Finance* **8: 1**: 93–124.

Piterbarg, V. (2004). Pricing and hedging callable libor exotics in forward libor models, *Journal of Computational Finance* **8: 2**: 65–117.

Rasmussen, N. (2005). Efficient control variates for monte-carlo valuation of american options, *Journal of Computational Finance* **9: 1**: 2–12.

Rebonato, R. (1998). *Interest-rate Option Models*, second edn, John Wiley & Sons Ltd.

Rebonato, R. (1999). *Volatility and Correlation*, John Wiley & Sons Ltd.

Rebonato, R. (2002). *Modern Pricing of Interest-rate Derivatives*, Princeton University Press.

Rogers, L. (2002). Monte carlo valuation of american options, *Mathematical Finance* **12**: 271–286.

Schoenmakers, J. and Coffey, B. (2003). Systematic generation of parametric correlation structures for the libor market model, *International Journal of Theoretical and Applied Finance* **6: 5**: 507–519.

Vasicek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics* **5**: 177–188.

Wilmott, P. (2006). *Paul Wilmott on Quantitative Finance*, second edn, John Wiley & Sons Ltd.

Appendix

A Financial calculus

Financial mathematics, or financial calculus, is the mathematical discipline concerning financial models. It contains modelling of the dynamics of financial assets and the pricing of derivatives whose value depends on these assets. Because these dynamics are typically stochastic, financial mathematics is based on stochastic calculus, especially stochastic differential equations. Since Black and Scholes (1973) developed a their famous option pricing formula, financial calculus has been a rapidly growing area of research. The derivation and application of the Libor market model utilizes several theorems and concepts from financial calculus.

This thesis does not provide an introduction to financial calculus. We will provide some useful references for anyone who is not familiar with it. Several books provide an introduction to financial calculus (e.g. Baxter and Rennie, 1996; Björk, 2004). Most of these books also provide an introduction to stochastic calculus (usually in the appendix, for more imforation, see e.g Brzezniak and Zastawniak, 1998). More information on (the valuation of) financial derivatives can be found in (Hull, 2003). For more information on interest rate derivatives and the Libor Market model in particular we suggest (Rebonato, 2002; Brigo and Mercurio, 2001).

A.1 Black's formula

As an exception to the remark above, we present Black's formula, because it is used very frequently throughout this thesis.

Theorem 1 (Black's formula) *Consider a European call option on a variable with value $V(t)$. Let T be the maturity time of the option, $F_T(t)$ be the forward price of V at time t with maturity T, X be the stike price and σ be the volatility of F_T. Assume that $V(T)$ is lognormally distributed with the expectation and volatility given by*

$$\ln(V(T)) \sim \mathcal{N}\left(\ln(F_T(t)), \sigma\sqrt{T-t}\right).$$

Under this assumption, the price of the call option at time t is given by:

$$c = P(t,T)\left[F_T(t)N(d_1) - XN(d_2)\right]$$

where

$$\begin{aligned} d_1 &= \frac{\ln(F_T(t)/X) + \sigma^2(T-t)/2}{\sigma\sqrt{T-t}} \\ d_2 &= \frac{\ln(F_T(t)/X) - \sigma^2(T-t)/2}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t} \end{aligned}$$

B Antithetic sampling for monotonic functions

Here we proof that $\operatorname{cov}(f^+ \equiv f(z), f(-z)) \leq 0$ when $f(z)$ is a monotonic convex function. This shows when antithetic sampling can be used.

For all x, y and two increasing functions f and g there holds:

$$(f(x) - f(y))\,(g(x) - g(y)) \geq 0$$

Now if X, Y are i.i.d. Taking the expectation we can write:

$$\begin{aligned}
(f(X) - f(Y))\,(g(X) - g(Y)) &\geq 0 \\
\mathbb{E}\left[(f(X) - f(Y))\,(g(X) - g(Y))\right] &\geq 0 \\
\mathbb{E}\left[f(X)g(X)\right] + \mathbb{E}\left[f(Y)g(Y)\right] &\geq \mathbb{E}\left[f(X)g(Y) + f(Y)g(X)\right] = \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(Y)\right] + \mathbb{E}\left[f(Y)\right]\mathbb{E}\left[g(X)\right] \\
\mathbb{E}\left[f(X)g(X)\right] &\geq \mathbb{E}\left[f(X)\right]\mathbb{E}\left[g(X)\right]
\end{aligned}$$

So $\operatorname{cov}(f(X), g(X)) \geq 0$. Now take $g(X) = -f(-X)$. if f is increasing, so is $\dot{g}$. Then it holds that

$$\operatorname{cov}(f(X), f(-X)) = -\operatorname{cov}(f(X), -f(-X)) \leq 0$$

C Parameters

Initial Libor curve and volatility parameters Φ_i:

i	L_i	Φ_i
1	0.023	-
2	0.025	0.153
3	0.027	0.143
4	0.027	0.140
5	0.031	0.140
6	0.031	0.139
7	0.033	0.138
8	0.034	0.137
9	0.036	0.136
10	0.036	0.135
11	0.038	0.134
12	0.039	0.132

Volatility and correlation parameters:

a	0.976
b	2.000
c	1.500
d	0.500
ρ_∞	0.663

D Alternative ways to estimate the barrier

In chapter 13 we presented two methods two find an optimal control variate. For the first method, the shifted cap where the barrier is estimated by a regression, there are several alternatives. Here we will discuss a dynamic extension. Because the results were mediocre, we have not discussed them in this thesis. But since it may be an interesting method for other CLE's, we discuss it here shortly. Next we will give other possible ways to extend the shifted cap method.

D.1 Dynamic shifted cap

In the method discussed in section 13.4, the barriers B_i are computed at T_0 and are constant over time. Therefore I will call this the static method. We can also try to update the barriers at each possible exercise date to get a dynamic strategy (similar to the dynamic method for the snowball in section 14.2.2). In that way it is possible to incorporate information about the current state of the Libors at future dates.

We can do this by not only regressing . $\min\limits_{1\leq j\leq i-1} H_{j+1}(T_j)$ but also $\min\limits_{k\leq j\leq i-1} H_{j+1}(T_j)$ for $2 \leq k \leq N$, to get the optimal barrier at a later date T_{k-1}. In this way we can improve the estimate for B_i at each exercise date, like we did in section 14.2.2. In addition, in the dynamic strategy, we can also take other state variables into account. We can add to the constant a_i a state dependent function. Suppose we have at time T_i, some vector of T_i-measurable state variables X_i, then we can try to improve the regression by

$$\min_{k\leq j\leq i-1} H_{j+1}(T_j) \approx a_i + c_i' X_i + b_i L_i(T_{i-1}) \qquad i \geq j+2$$

The barrier is then given by $B_i = -\frac{a_i + c_i' X_j}{b_i}$. In this way the barrier can be made path-dependent.

D.2 Other ideas

Further improvements could be:

- Replace $1_{\{L_i > B_i\}}$ by $1_{\left\{\max_{j\leq i}(L_j) > B_i\right\}}$. So the payoff occurs when the Libor rate hits a barrier on one of the tenor dates. This is very similar to a discrete barrier option. We cannot value discrete barrier options exactly, but very good approximations exist.
- Replace $1_{\{L_i > B_i\}}$ by $1_{\{X(T_i) > B_i\}}$ where X is another variable. This may help to get a better approximation of the exercise strategy. For example, it could be the swap rate, because the exercise strategy clearly depends stronger on the swap rate compared to the single Libor rate L_i. Unfortunately, the valuation is much harder, because the value now depends on two (usually correlated) underlying variables.
- Use weighted least squares, where the weights are proportional to the inverse of $\min\limits_{1\leq j\leq i-1} H_{j+1}(T_j)$ and/or to $L_i - R$, to focus on those paths where the correct estimate is more important. We tried to use this method, but the results were disappointing.
- Do not use a regression at all, but estimate the barrier in another way. We have tried to do some kind of brute-force optimization by first optimizing the correlation over the first cash flow date by choosing B_1. Next we looked at the correlation for the (discounted) sum of the first two cash flows, by changing B_2, etc. Unfortunately, even this did not yield better results compared to the regression method.