An Implementation of
the Displaced Diffusion, Stochastic Volatility Extension
of
the LIBOR Market Model

A Comparison to the Standard Model

Handelshøjskolen, Aarhus Universitet
July 2007
1 Introduction .................................................................................................................................3
  1.1 Problem statement ...........................................................................................................3
  1.2 Delimitation ....................................................................................................................4
  1.3 Structure of the paper ....................................................................................................5

2 Notation and definitions ........................................................................................................6
  2.1 The interest rate cap .....................................................................................................8
  2.2 Exotic caps ..................................................................................................................10

3 Change of numeraire ............................................................................................................13
  3.1 Relative prices ..........................................................................................................13
  3.2 Martingales ................................................................................................................13
  3.3 Changing numeraire ................................................................................................14
  3.4 The Girsanov theorem ............................................................................................17

4 The Black formula ..............................................................................................................19

5 The standard LIBOR market model ..................................................................................22
  5.1 No arbitrage drifts ....................................................................................................24

6 Volatility ............................................................................................................................27
  6.1 Desirable properties of the volatility specification ....................................................27
  6.2 Model and empirical implied volatilities ..................................................................28
  6.3 Different approaches to obtain a non constant volatility ........................................31
    6.3.1 Constant elasticity of variance models ..........................................................32
    6.3.2 Displaced diffusion models ...........................................................................33
    6.3.3 Jump diffusion models ..................................................................................34
    6.3.4 Regime switching models .............................................................................34
    6.3.5 Stochastic volatility models .........................................................................35

7 The extended LIBOR market model ..................................................................................37
  7.1 No arbitrage drifts ....................................................................................................39
  7.2 Comparison of the models .........................................................................................40

8 Implementation issues ........................................................................................................41
  8.1 Choosing the measure/numeraire .............................................................................41
  8.2 Sampling the forward rates distribution ....................................................................42
    8.2.1 The covariance matrix ....................................................................................44
8.2.1.1 Low discrepancy numbers ...............................................................47
8.2.2 Conditionally deterministic drift terms ..............................................49
8.3 Payoff and expected value ....................................................................53
9 Data ...........................................................................................................55
10 Calibration ...............................................................................................60
  10.1 The yield curve ..................................................................................60
  10.1.1 Nelson-Siegel ................................................................................61
  10.1.2 Estimation .....................................................................................61
  10.2 Correlation .......................................................................................63
  10.3 Volatility ...........................................................................................65
  10.3.1 Calibration of the standard model ..................................................65
  10.3.2 Calibration of the extended model ..................................................70
11 Convergence ............................................................................................81
  11.1 Convergence of calibration setup ......................................................84
  11.2 Convergence of pricing setup .............................................................86
12 Pricing .....................................................................................................90
13 Concluding remarks ..............................................................................100
  13.1 Conclusion .......................................................................................100
  13.2 Appraisal of the extended model .......................................................102
  13.3 Suggestions for further work ............................................................103

The introduction makes up for the mandatory abstract.
1 Introduction

The LIBOR market model attributed mainly to Brace, Gatarek & Musiela (1997) and Jamshidian (1997) is an interest rate model, which models directly a set of non-overlapping discretely compounded forward rates. It has become increasingly popular as it allows for a flexible framework for pricing of a variety of interest rate products. The dynamics of the model results in log normally distributed forward rates, which is the main assumption of the Black formula. This means that the LIBOR market model can recover exactly the caplet prices computed by the Black formula, additionally it affords a financial interpretation of the Black implied volatility.

This has increased the acceptance of the model as the Black formula is now the market standard.

However given log normally distributed forward rates the model can only produce implied volatilities that are constant across strikes. As we will show that the implied volatilities observed in the market are not constant across strikes the forward rates cannot be log normally distributed.

Researchers and practitioners have therefore searched for extensions of the LIBOR market model that could remedy this deficiency of the standard model. Several models have been developed that are able to reproduce a non-flat volatility. But no consensus has been reached on the best model.

1.1 Problem statement

In this paper we will analyze one of these extensions proposed by Joshi & Rebonato (2003): The displaced diffusion, stochastic volatility extension of the LIBOR market model. The model has two distinct features to induce a non-flat volatility. The displaced diffusion feature results in a skew in implied volatilities and the stochastic volatility results in a smile.

The paper will analyze the differences encountered when going from the standard to the extended LIBOR market model. The analysis will concentrate on a theoretical comparison of the stochastic differential equations assumed for the forward rates in the two models, a comparison of the calibration of the models both in terms of procedure and fit to market data and an analysis of the differences concerned with
pricing, both implementation issues and differences in the prices produced by the models.

As the extension consists of two separable features we will to a large extend treat the two features independently, as we will also analyse the calibration fit and pricing differences for each feature both separable and in combination.

We will show that the model is an intuitive extension of the standard model, which allows a financial interpretation of the parameters. The model affords a considerably better fit to market data than the standard model, but this improvement in fit comes at the cost of a substantial increase in calibration and computation time, due to the need for repeated samplings of the stochastic volatility. We will also show that the models result in differing prices and the analysis will show that the two separable features of the model are often acting in opposing direction and that the pricing differences will differ for changing characteristics of the products.

1.2 Delimitation

The empirical part of the analysis is conducted based on one trading day, as an analysis across different trading days would extent the scope of the paper considerably.

The issue of day count conventions has not been addressed fully in the analysis. We assume the 30/360 day count convention for both the products we calibrate to and the products we price. We are aware that the products that we use for calibration of the models are not quoted with this particular day count convention, but a preliminary analysis has showed that the assumption has only little influence, and as the scope of the analysis is a general comparison of the models, we are comfortable with the assumption.

In addition to these delimitations some of the issues we address, including estimation of the correlation structure and bootstrapping caplet volatilities from cap volatilities, could each be investigate in greater depth. But in order to maintain focus on the comparison of the models we restrict our treatment of these necessary but secondary issues to the point where we obtain a working solution. But in a practical implementation of the models these would be issues to be addressed further. We asses that this delimitation has a very small impact on the comparison as the models are based on the same methods.
1.3 Structure of the paper


The paper is structured in three parts. The first part establishes the foundation for the following parts by introducing notation and definitions including the products that we consider either as calibration instruments or in the pricing comparison. This part also includes a brief section introducing the mathematical concepts and results employed later in the paper. Finally the first part includes a presentation of the Black formula, as many concepts are directly linked to this formula and the formula is utilized in the calibration.

The second part is devoted to an analysis of the models. First we establish the standard model and the connection to the Black formula. We then turn to an analysis of the volatility, comparing the volatility surface observed in the market to the volatility surface obtainable from the standard model. Before turning to the displaced diffusion, stochastic volatility model we give a brief overview of some of the possible model classes which introduce non-flat volatility. Having established the models we address the implementation issues concerned with the models.

The last part is an empirical part containing the results of the implementation. First we will present the market data that are used in the implementation. Then we describe the calibration procedures used to obtain the model parameters of the yield curve, the correlation and the volatility. Next we address the issue of convergence of the model, and finally we analyze the pricing differences between the models.

We end the paper with a conclusion, an appraisal of the extension and suggestions for further research.
2 Notation and definitions

To facilitate the analysis the following section will contain central notation and definitions. We will also introduce products which are fundamental to the analysis.

First we consider some concepts linked to the term structure of interest rates. The term structure of interest rates is a representation of interest rates as a function of time to maturity.

The tenor structure defines the possible maturity times as $T_1, T_2, \ldots, T_N$ and the tenor is defined as $\tau_i = T_{i+1} - T_i$.

The discount rates, meaning the price at time $t$ of a zero-coupon bond paying 1 unit of currency at maturity time, $T_i$, is denoted by $P(t, T_i)$. The spot interest rate at time $t$ for maturity $T_i$ is denoted by $R(t, T_i)$.

The discount rate and the spot interest rate are related by

$$P(t, T_i) = \frac{1}{\left(1 + \frac{R(t, T_i)}{m}\right)^{\tau_i m}}$$

(1)

where

$m$ is the compounding frequency of the spot rate

For semi-annual compounding the compounding frequency would be 2.

We also consider a set of spanning forward rates\(^1\). Given a set of zero-coupon bonds, we define forward rates as given by

$$f(t, T_i, \tau_i) = \frac{P(t, T_i) / P(t, T_i + \tau_i) - 1}{\tau_i}$$

(2)

\(^1\) Spanning forward rates will refer to a set of consecutive non-overlapping forward rates
It is the interest rate that we can contract for at time $t$ for a loan between future time $T_i$ and $T_{i+1}$. $T_i$ and $T_{i+1}$ are called the reset and the maturity of the $i$’th forward rate. The concepts of reset and maturity can be illustrated by seeing that

$$f(T_i, T_{i+1}, \tau_i) = R(T_i, T_{i+1} + \tau_i)$$

So the forward rate for a given period can change as time passes, but at the reset time, it is fixed (resets) and equal to the spot rate for the same time period. The interest is due for payment at the end of the period, i.e. at the maturity time of the forward rate.

As we work with the 30/360 day count convention the tenor will be constant at 0,5 year, and we will often suppress the dependence on tenor and write $f(t, T_i)$ or equivalently (but lighter) $f_i(t)$.

The forward rates will be expressed with a compounding frequency equal to the tenor, i.e. semiannual compounding.

The timing convention and the relations between the forward rates, discount rates and spot rates can be illustrated by the following figure.

**Figure 1:** Illustration of timing conventions for forward rates, discount rates and spot rates

The LIBOR market model models the dynamics of the set of forward rates, i.e. describes $f(t, T_i, \tau)$ as time elapses, keeping $T_i$ and $\tau$ fixed. So what we model for each forward rate is the interest rate for a given future period of time. As we have seen these forward rates determine the short spot rate at their reset, and we have thereby modelled the future short spot rates. But at any time we could also extract the current
values of the forward rates that has not yet reset, and from them calculate the whole yield curve by inverting Equation (2). Consider for example a product whose payoff is determined in one years time as some function of the four year spot rate. We would then model the forward rates for the next five years. These rates are evolved one year to the time where the payoff is determined. At this time we can calculate the four year spot rate from the forward rate that has just reset (the one period spot rate) and the ones that has not yet reset.

This shows that the LIBOR market model can be used for pricing of a variety of interest rate products.

At a later time we will also need to be able to calculate the swap rate.

The equilibrium swap rate for an interest rate swap maturing at \( t_{n+1} \), can be shown to be given by:

\[
R_s = \frac{1 - P(0, t_{n+1})}{\tau \sum_{i=1}^{n} P(0, t_{i+1})}
\]

### 2.1 The interest rate cap

The LIBOR market model is calibrated to a set of interest rate caps. An interest rate cap is a collection of options, caplets, each paying if some future interest rate is above a certain level, making it a suitable instrument as insurance against increasing interest rates for a borrower paying floating interest rate on his debt.

The properties of the caplet can be illustrated by an example: A borrower pays a floating rate on his debt as the interest rate for the period \( T_{i-\tau} \) to \( T_i \) is determined at time \( T_{i-\tau} \) as \( \tau \)-spot EURIBOR with compounding equal to \( \tau \). The interest rate payment must be paid at \( T_i \). His interest payment at time \( T_i \) assuming no amortization of the debt is therefore

\[
L \times \tau \times R(T_i - \tau, \tau)
\]

where

\( L \) is the principal of the debt
He is therefore concerned about increasing interest rates as this would lead to an increase in his future interest payment.

He could therefore buy an interest caplet for the period, $T_i - \tau$ to $T_i$. The caplet will return a payoff if the spot EURIBOR at time $T_i - \tau$ is greater than the cap rate/strike, $R_K$. The payoff will be paid at time $T_i$. The payoff at time $T_i$ from the caplet is given by:

$$NP \times \tau \times \max(R(T_i - \tau, \tau) - R_K, 0)$$

where

(6) $NP$ is the notional principal contracted for $R(T_i - \tau, \tau)$ is the spot $\tau$ - period EURIBOR one period before maturity $R_K$ is the cap rate Both rates expressed with a compounding frequency equal to the tenor

The holder of the caplet thus receives an interest payment on the notional principal equal to the positive difference between the cap rate and the spot rate. The cap can therefore be seen as a call option on the forward rate resetting at $T_i - \tau$, and it puts an upper bound on the net interest rate paid. In alignment with the timing standard for the forward rates illustrated above, the size of the payoff is determined as the forward rate resets at $T_i - \tau$, while the amount is received one period later at the time of maturity, $T_i$. The caplet only provides insurance for this particular interest payment. Instead the borrower could purchase a cap, which is a collection of caplets covering the period from now to maturity.

The price at time $t$ of a cap maturing at time $T_n$ is given by

$$C(t, T_n) = \sum_{i=2}^{n} c(t, T_i)$$

(7) where $c(t, T_i)$ is the value of the caplet maturing at $T_i$

We see that the first caplet matures at $T_2$ and the last at the maturity of the cap. There is no caplet maturing at $T_1$ as the interest rate underlying this cap has already reset.
A 10 year semi-annual cap is thereby made up by 9 consecutive caplets, the first paying in 1 years time based on the forward rate resetting in half a year, the last paying in 10 years time based on the forward rate resetting in 9,5 year.

Each caplet is said to be in-the-money (ITM), at-the-money (ATM) or out-of-the-money (OTM) if the underlying forward rate is currently higher, equal to or lower than the cap rate respectively.

A cap is defined as ATM for a cap rate equal to the swap rate for a swap that has pay offs at the same time as the cap.\(^2\)

### 2.2 Exotic caps

The pricing differences between the standard and the extended LIBOR market model will be analysed in the context of two exotic caps: The Ratchet and the Sticky cap.\(^3\)

Both products are path dependent as the payoffs of the underlying caplets are not only a function of the forward rate maturing at the maturity time of the caplet, but also a function of the forward rates that has already reset.

For both products the payoff function in Equation (6) is basically unaltered, but the strike of the underlying caplets are made path dependent. Defining the vector of strikes as \( R_k \), with elements \( R_k^i \) being the strike of the i’th caplet, the strikes are given by:

\[
\begin{align*}
R_k^i &= g(f) \\
\end{align*}
\]

The strikes are a function of all forward rates.

The reason that path dependent caps are a interesting products in a comparison of the standard and the extended model is that they depend on the correlation between forward rates. As will be shown later correlation between forward rates will depend on the volatility of the forward rates, and it is therefore expected that going from the

\(^2\) I.e. a forward swap with the same time of maturity and tenor as the cap. The time to the start of the swap equal to the tenor.

\(^3\) Hull (2003)
standard to the extended model will affect the price and hedging parameters of Ratchet caps.

For the Ratchet cap the strike is given by

\[ R'_k = f_{i-1}(T_{i-1}) + c \]

(9)

where

c is a constant margin

For each caplet the strike is given by the forward rate as it has reset one period before the forward rate underlying the caplet plus a constant spread. For a borrower paying floating interest rate the Ratchet cap will result in the situation, that the borrower will always know his maximum interest payment at the next date of payment, but all subsequent payments will be unknown.

For the Sticky cap the strike is given by

\[ R'^i_k = \max(f_{i-1}(T_{i-1}), R'^{i-1}_k) + c \]

(10)

where

c is a constant margin

For each caplet the strike is given by the minimum of the forward rate as it has reset one period before and the cap rate one period before, i.e. the capped interest rate for the period before, plus a constant spread. The sticky cap will therefore result in a situation where the maximum payment for the next period will be the payment of the current period plus a spread.

The product characteristics can be illustrated by the following graphs showing the evolution of the cap rate in the two products for the same forward rate evolution.
Figure 2: Realized cap rates for Ratchet and Sticky cap for a given forward rate path

The graphs show that the cap rates evolve identically when the interest rate is falling or rising slightly, but when the interest rate rises more than slightly the increase in the cap rate from period to period is restricted for the Sticky, as the cap rate can only go up by the margin. For the Ratchet the cap rate is always set at the value of the previous forward rate plus the margin. Thereby the Ratchet allows larger jumps in the cap rate than the Sticky.

From this it is evident that the Sticky cap must be more valuable than the Ratchet for the same margin.
3 Change of numeraire

When developing the Libor market model the procedure of changing numeraire/changing measure is employed. A brief explanation of the procedure and the concepts of numeraire and measure is therefore warranted at this stage. The goal of this section is to explain the concepts in an intuitive matter, which exactly enables the understanding of the subsequent sections. In obtaining this goal no attempts will be made to present the full and rigorous implications of the concepts. For a full presentation of the concepts see e.g. Neftci (2000).

3.1 Relative prices

The procedure of changing numeraire is closely linked to the concept of relative prices. Given the price, \( P_i \), of asset \( i \) and the price, \( N \), of another traded asset the relative price of asset \( i \) with respect to the other asset is defined as

\[
Z_i = \frac{P_i}{N}
\]

The relative price is thereby the price of an asset expressed as units of the other asset, the numeraire. Any asset with a strictly positive price in any state of the world can be used as numeraire.\(^4\) It is apparent that the modelling of \( Z_i \) is equivalent to modelling \( P_i \). But as will be shown later, by choosing carefully the asset used as numeraire one will obtain nice modelling properties.

3.2 Martingales

The next concept to be introduced is a martingale. Given a probability space \( (\Omega, \mathcal{F}, Q) \)\(^5\) the stochastic process \( X \) is a martingale under \( Q \) if it satisfies

\[
E^Q(X_T | \mathcal{F}_t) = X_t, \quad T > t
\]

---

\(^4\) Rebonato (1998)

\(^5\) For a description of the probability space concept see Neftci (2000)
The time t expectation under the probability measure $Q_t$ of the value of the process at some later time $T$ is equal to the value of the process at time $t$, i.e. the process is driftless.

To simplify notation we will denote the time t expectation by $E_t^{Q_t}(X_T)$ in the following.

### 3.3 Changing numeraire

The probability measure assigns probabilities to the possible states of the world. In the real world the probability measure is given by $Q$, and a process that is a martingale under $Q$ would be a martingale in the physical world. But we could also consider probability measures different from the physical one. These measures would describe probabilities in these worlds. For instance in the well known risk free world the probabilities are given by the risk free measure $Q_0$.

In the absence of arbitrage it is possible to find for each chosen numeraire an equivalent probability measure, under which relative prices are martingales; this measure is called the equivalent martingale measure. Changing numeraire/changing measure is the procedure of going from one numeraire to another and changing the measure so that the relative prices are martingales.

The mathematical representation of this reveals a powerful result. Defining $Q_N$ as the equivalent martingale associated with the numeraire $N$ any asset pricing problem can be stated as the following with $V(t)$ denoting the time $t$ value of the asset.

\[
\frac{V(t)}{N(t)} = E_t^{Q_N}\left(\frac{V(T)}{N(T)}\right), \quad T > t
\]

(13)

\[
V(t) = N(t)E_t^{Q_0}\left(\frac{V(T)}{N(T)}\right)
\]

It shows that the price of any asset can be expressed as the expected relative price of the asset under the equivalent martingale measure associated with the numeraire times the current price of the numeraire.
In option pricing we know the option value at maturity given the state of the world through the payoff specification. As we often take the current value of the numeraire as given the option pricing problem therefore becomes a problem of evaluating the expectation under the equivalent martingale measure of the payoff relative to the future value of the numeraire. The numeraire should therefore be chosen in order to simplify the evaluation of the expectation as much as possible.

An example of the concept can be cast in the framework of the Black-Scholes model. Black & Scholes (1973) discovered that the pricing of a European call option could be greatly simplified if one chooses as numeraire the rolled up money market account. The rolled up money market account is the asset corresponding to investing 1 unit of currency at time 0 at the prevailing instantaneous spot rate $r(t)$ and rolling the investment over, always at the instantaneous spot rate. The value of the account at time $t$ is given by

$$B(t) = \exp\left[ \int_0^t r(s) \, ds \right]$$

Under this numeraire the equivalent martingale measure is the risk free measure, $Q_0$, under which all assets grow at the risk free rate.

Inserting this in Equation (13) gives us the following expression for the value of the call option at time 0, $C(0)$

$$C(0) = B(0)E_0^Q \left( \frac{C(T)}{B(T)} \right)$$

The second equality is obtained using $B(0) = 1$ and the last is obtained by the fact that in the Black-Scholes world interest rates are deterministic and can therefore be taken outside the expectation. The choice of numeraire simplifies the evaluation of the
expectation as the dynamics of the underlying under the risk free measure results in a closed form solution of the expectation. Inserting the closed form solution in Equation (15) provides us with the Black-Scholes formula.

The rolled up money market account is thereby a very convenient numeraire given the assumptions of Black and Scholes as it yields a closed form solution for the call option.

The use of the rolled up money market account is especially useful in the Black-Scholes world as they assume deterministic interest rates. When valuing interest rate products we do not assume deterministic interest rates, and the attractiveness of the rolled up money market account is reduced. Instead we consider the terminal measure,\(^6\) which is the martingale measure corresponding to choosing as numeraire the zero coupon bond maturing at the same time as the asset under consideration. Since the price of the numeraire is 1 at maturity and denoting the terminal measure corresponding to the zero coupon bond maturing at \(T_n\) by \(Q_n\), this leads to the following representation of the general pricing formula in Equation (13)

\[
V(t) = P(t, T_n) E^Q_n \left( V(T_n) \right)
\]

The terminal measure can also be defined in terms of a forward rate as the martingale measure using as numeraire the zero coupon bond maturing at the maturity of the forward rate. This highlights that opposed to the risk free measure the terminal measure depends on the maturity of the forward rate. Sometimes we will use the statement “under the terminal measure” for a collection of forward rates. By this we mean “under the terminal measures of each forward rate.”

Under the terminal measure all relative prices are martingales. But as the forward rates are not traded assets, we cannot conclude that they are martingales. However it can be shown\(^7\) though that under the terminal measure each forward rate is a martingale and that a forward rate will not be a martingale under the terminal measure

\(^6\) Also called the forward measure

\(^7\) Rebonato (2002)
of another forward rate; for a given terminal measure one and only one forward rate will be a martingale.

A central result can be obtained by combining the above with the result that any diffusion with deterministic drift and volatility leads to a log normal distribution. Under the forward measure the forward rates have zero drifts and therefore the forward rates will be log normally distributed under the forward measure as long as the volatility is deterministic. Additionally it can be proved that under a specific measure only one forward rate can be log normally distributed. We therefore have that under a certain terminal measure, only the forward rate that pays off at the maturity of the numeraire will be log normally distributed, any other forward rate will not.

### 3.4 The Girsanov theorem

Given that we can choose different numeraires and thereby change the measure we need to examine how the dynamics of a process change when we go from one measure to another.

The Girsanov theorem states that if $W_t$ is a P-brownian motion, i.e. a Brownian motion under P, and Q is a probability measure equivalent to P, then there exists a process $\gamma_t$ such that

$$\tilde{W}_t = W_t + \int_0^t \gamma_s \, ds$$

is a Q-brownian motion. That is, under Q the P-brownian motion plus a drift is a Brownian motion. We see that the consequence of changing measure is that the drift of the process changes; under P the process $\tilde{W}_t$ has a drift, under Q it does not. Obviously the change of measure can also convert a non-drifting process under P to a drifting process under Q. An important implication of Girsanov’s theorem is that the change of measure does not change the volatility of a process. The process

---

8 Rebonato (2002)

9 And the forward rates follow a diffusion

10 It assigns zero probabilities to the same states of the world
$\tilde{W}_t$ has the same volatility under both $P$ and $Q$. The same carries over to the correlation between two processes, i.e. when changing measure the processes under consideration are transformed in a way that changes the drifts of the processes while the volatility and correlations remain the same.
4 The Black formula

The option pricing formula developed by Black and Scholes in their seminal article (Black & Scholes 1973) was originally developed for the pricing of options under the assumption of constant interest rate and volatility. But the use of the formula was soon expanded to include non-constant (but still deterministic) interest rates and volatilities and options on other different assets. One of the developments was achieved by Black himself (Black 1976) by the derivation of a formula for pricing options written on commodities. To circumvent the predictability of the spot price of commodities Black modeled the options as written on the future price of the commodity rather than the spot price as is the case for the Black-Scholes formula.

As interest rates are also predictable, a version of the Black formula was soon formulated to price interest rate caplets. The main assumption underlying the derivation of the formula is that forward rates are log normally distributed with a variance of $\sigma_{Black}^2(T)$ at maturity, $T$.

The Black formula for the price of a caplet with tenor, $\tau$, and maturing at $T+\tau$ writes:

$$
\begin{align*}
  c_i^{Black} &= P(t,T + \tau) \left[ f(t,T,T + \tau) N(d_1) - R_K N(d_2) \right] NP \tau \\
  d_1 &= \frac{\ln\left[ f(t,T,T + \tau) / R_K \right] + \sigma_{Black}^2 (T - t) / 2}{\sigma_{Black} \sqrt{T - t}} \\
  d_2 &= d_1 - \sigma_{Black} \sqrt{T - t}
\end{align*}
$$

(18)

where

- $N(x)$ is the standard cumulative normal distribution
- $R_K$ is the cap rate
- $\sigma_{Black}$ is the Black volatility

All the parameters and variables of the formula but the volatility are directly observable in the market. The volatility, $\sigma_{Black}$, is not directly observable, and the market practice is to quote the prices of caplets as the implied volatility of the option, i.e. the number that put into the formula as $\sigma_{Black}$ will give the market price of the

---

11 Due to e.g. harvest periods in the case of agricultural commodities
caplet. As the market quotes prices for a range of maturities and cap rates we can obtain a matrix of implied volatilities, and the implied volatilities are identified by \( \sigma_{Black}(T_i, R_{K,j}) \) where \( T_i \) identifies the reset time of the underlying forward rate and \( R_{K,j} \) identifies the cap rate. When plotted in cap rate/reset time space we obtain the implied volatility surface. If we consider only one cap rate the implied volatility as a function of maturity describes the term structure of volatility.

As we have stated before, a set of forward rates cannot be simultaneously log-normal, it would therefore seem that the Black formula is internally inconsistent if used for pricing caplets at different maturities, as it assumes log normality of the forward rates. But if we add the qualifier that each forward rate should be log-normally distributed under its own terminal measure, and the pricing is done in each of these measures, the assumption of log-normality will be recovered. And as the pricing is independent of the measure, we will recover the same price.

However powerful the Black formula is, it has two major drawbacks. First of all it does not provide us with the dynamics of the forward rates. The formula only assumes that each forward rate is log normally distributed under its own terminal measure. This assumption is sufficient for the pricing of caplets as they only depend on one forward rate. This means that the payoff depends only on the distribution of the forward rate at maturity, the possible paths leading to this distribution is irrelevant, and also it allows us to consider each forward rate individually under its own terminal measure.

The value of exotic products could depend on multiple forward rates at points in time different from the maturity. This means that we need a setup in which we can obtain the joint distributions of multiple forward rates under one common measure and we need marginal distributions not just the terminal ones. In order to obtain this we need the dynamics of the forward rates.

A variety of dynamics would result in log normal forward rates and thereby retain the validity of the Black formula. We know from Section 3.3 that among others any diffusion with deterministic drift and volatility will produce a log normal distribution.
Second under the assumption of log normality the Black implied volatility is linked directly to the variance of the forward rate at maturity. This means that caplets of the same maturity but with different strikes should exhibit the same Black volatility as the same forward rate is underlying the caplets. So when we assume log normal forward rates we also assume a flat volatility surface across strikes. As we will see later this is not the case. Caplets differing in strike will also differ in Black implied volatility. This is evidence against the assumption of log normality of forward rates, and dynamics leading to log normality would not be consistent with the market.
5 The standard LIBOR market model

The standard LIBOR market model resolves the first deficiency of the Black formula by providing dynamics for a set of forward rates under one common measure.

The forward rates are modelled as\(^{12}\)

\[
\frac{df(t)}{f(t)} = \mu(f, t)dt + \sigma(t)dz(t)
\]

where

\[
\frac{df}{f}
\]

is a \(n \times 1\) vector of percentage increments of forward rates

\[\mu(f, t)\]

is a \(n \times 1\) vector of drift terms, which depend on both \(f\) and \(t\)

\[dz(t)\]

is a \(n \times 1\) vector of correlated Brownian motions under the probability measure \(Q\)

\[\sigma(t)\]

is a \(n \times n\) diagonal matrix with the \((i, i)\)'th element, \(\sigma_i(t)\)

being the instantaneous volatilities of the \(i\)'th forward rates

The correlation between the forward rates are given by the correlation matrix, \(\Gamma(t)\), with elements:

\[\rho_{ij} = E\left[ dz_i(t)dz_j(t) \right]\]

Sometimes it will be more illustrative to work with the dynamics of a single forward rate in stead of the full matrix formulation above. For the \(i\)'th forward rate the model reads:

\[
\frac{df_i(t)}{f_i(t)} = \mu_i(f, t)dt + \sigma_i(t)dz_i(t)
\]

---

\(^{12}\) The model specification is from Rebonato (2002)
Integration leads to the following expression for the forward rate at some later time, $T$.

\[
    f_i(T) = f_i(t) \exp \left( \int_t^T \left[ \mu_i(f_i(u), u) - \frac{1}{2} \sigma_i^2(u) \right] du + \int_t^T \sigma_i \, dz_i(u) \, du \right)
\]

This representation will prove useful in some of the later analysis. But for now it suffices to say that the equation highlights that evolving the forward rates narrows down to evaluation of two integrals; a drift integral and a volatility integral.

By choosing these dynamics some modelling choices have been made. First of all the dynamics of the forward rates is described by a pure diffusion model and all uncertainty is described by a number of Brownian motions. This rules out the possibility of jumps in the process. Second by choosing to model the forward rates as geometric Brownian motions the model insures that the forward rates cannot take negative values, and also the volatility of the percentage increments are constant regardless of the level of the forward rates.

The standard model restricts the volatility to be deterministic. It does not necessarily need to be constant, but it must only be a function of time. As the specification of the volatility is a key issue of the model, and some additional analysis is necessary for this subject a full section, following the current, is devoted to volatility.

The main reason for the popularity of the LIBOR market model is the recovery of Black prices. The model is capable of exactly reproducing the Black prices of caplets. The caplet value depends on one forward rate only and each caplet can be considered in isolation. For each caplet we can therefore change the measure to the terminal measure. Under this measure the forward rate will be a martingale and therefore the dynamics reduce to a driftless diffusion with deterministic volatility.

By invoking the result from Section 3.3 that any diffusion with deterministic drift and volatility results in a log normal distribution we see that the dynamics of the standard model produce log normally distributed forward rates. As the LIBOR market model

\[\text{Research exploring the possibility of jumps are being conducted see e.g. Glasserman & Kou (2003)}\]
implicitly assumes the same distribution of the underlying the LIBOR market model will produce the same prices as the Black formula.

The distribution under the terminal measure of the forward rates in the LIBOR market model in terms of the volatility is given by

\[
\ln f_i(T) \sim N\left(0, \int_t^T \sigma_i^2(u)du\right)
\]

The variance of the forward rate at maturity is the integrated squared instantaneous volatility to the reset of the forward rate. As the variance at maturity in the Black formula is \(\sigma_{\text{Black}}^2(T_i)T_i\), the standard LIBOR market model can recover the Black price of any caplet as long as one chooses;

\[
\sigma_{\text{Black}}^2(T_i, R_{K,i})T_i = \int_t^{T_i} \sigma_i^2(u)du
\]

We see that the Standard model results in log normal forward rates and thereby the model can only produce a flat volatility surface across strikes.

### 5.1 No arbitrage drifts

We have seen that the forward rates are driftless under the terminal measure. But when using the standard LIBOR market model we will choose one measure\(^{14}\) and consider all the forward rates under this measure. As a result of the Girsanov theorem changing the measure from the forward measure to one common measure for all the forward rates will change the drifts of the forward rates processes while leaving unaltered the volatility and correlation. This means that the change of measure will not affect the volatility or correlation and it can therefore by obtained in one measure and applied directly in another. But the drifts will depend on the measure and only the forward rate maturing at the same time as the numeraire will be driftless. We will now show the drifts for the remaining forward rates.

---

\(^{14}\) The optimal choice of measure is discussed in section 8.1
As stated in Section 3.3 a necessary and sufficient condition of no arbitrage is that the relative prices of all traded asset should be martingales under the equivalent martingale measure associated with the chosen numeraire. As forward rates are not traded assets we cannot utilize this result directly for the forward rates. But an application of the result and some stochastic calculus\(^{15}\) leads to the following drift terms necessary to prevent arbitrage between the model forward rates when choosing as numeraire the zero-coupon bond maturing at the payoff time of forward rate \(f_j\), i.e. \(P(t,T_{j+1})\)

\[
\begin{align*}
\mu_i &= \sigma_i(t) \sum_{k=j+1}^{i} \frac{\sigma_k(t) \rho_{ik}(t) f_k(t) \tau_k}{1 + f_k(t) \tau_k} & \text{for } i > j \\
\mu_i &= -\sigma_i(t) \sum_{k=j+1}^{i} \frac{\sigma_k(t) \rho_{ik}(t) f_k(t) \tau_k}{1 + f_k(t) \tau_k} & \text{for } i < j \\
\mu_i &= 0 & \text{for } i = j
\end{align*}
\]  

(25)

As expected the forward rate which pays at the same time as the chosen numeraire, is driftless, i.e. a martingale. The other forward rates has non zero drifts. These drifts ensure that no arbitrage opportunities exist between forward rates under the equivalent martingale measure associated with the choice of numeraire.

We see that the drifts can be expressed as functions of instantaneous volatilities, correlations and forward rates, which are all quantities that we are already using in the diffusion term.

Additionally, as the expressions contain the forward rates, the drifts are state dependent. By looking at the expression for a forward rate at some later time using as numeraire, \(P(t,j+1)\), which matures after the maturity of the forward rate, we see why this constitutes a problem.

\[
\begin{align*}
f_i(T) &= f_i(t) \exp \left( \int_t^T \left[ -\sigma_i(u) \sum_{k=j+1}^{i} \frac{\sigma_k(u) \rho_{ik}(u) f_k(u) \tau_k}{1 + f_k(u) \tau_k} - \frac{1}{2} \sigma_i^2(u) \right] du + \int_t^T \sigma_i(u) d\xi_i(u) du \right)
\end{align*}
\]  

(26)

\(^{15}\) For a derivation see Appendix 1: No arbitrage drifts
The state dependency means that the forward rates are no longer log normally distributed, and the sampling is made much more difficult. The problem arises as the integral stemming from the state dependent drifts must be evaluated numerically. In a Monte Carlo simulation, this means that a fine stepped algorithm is necessary in order to evaluate and update the drift term frequently enough as the forward rate changes.

As it turns out different approximations (some very accurate) exist, so that we can approximate the state dependent drifts by a deterministic one, thereby recovering the log normality of the forward rates. This will be explored in detail in Section 8.2.2.
6 Volatility

Specification of the volatility in the Libor market model is a key issue which gives rise to some conceptual discussions. We have therefore devoted an entire section to this subject.

The section will start by discussing the properties that we would like the term structure of volatility from the model to exhibit. This discussion leads to the selection of a parameterization of the volatility.

Next the section will turn to a comparison between the different shapes of the implied volatility surfaces obtainable from the standard Libor Market Model and the empirically observed.

Finally the section will discuss different approaches to generate a non-constant volatility surface.

6.1 Desirable properties of the volatility specification

As Rebonato (2002) illustrates, the most typical shape of the term structure of volatility, is a hump shape.

Going from short to longer maturities the implied volatility is first increasing until it reaches a maximum around 2-3 years after which it decreases.

Although there are variations in the implied volatility, the implied volatility exhibits the same qualitative shape.

The first implication of this is that the volatility specification must be able to generate the humped shape. Second as long as we only consider deterministic volatility and given that we have no view on the future volatility we should consider time-homogenous volatility specifications. The term time-homogenous describes the property that the term structure of volatility is constant over time, meaning that the volatility of a 4 year caplet in 2 years time is the same as the volatility of a 2 year caplet today. Mathematically:

\[
\int_{T_i}^{T_h} \sigma_i(u)^2 \, du = \int_{T_{i+h}}^{T_i} \sigma_{i+k}(u)^2 \, du \text{ for all } i, k \text{ and } h. \quad h < i
\]
The interpretation of the "for all i, k and h" sentence is that the time homogeneity i.e. the same volatility for the same remaining maturity applies for any pair of caplets (i and k) and for any length of remaining time to maturity (h). The time homogeneity condition translates into a requirement that the volatility is a function of residual maturity (T-t) only.

Rebonato (1999) proposes a now widespread formulation of the instantaneous volatility function:

\[ \sigma_i(t) = \left[ a + b(T_i - t) \right] e^{-c(T_i - t)} + d \]

(28)

where

a, b, c and d are parameters to be determined

As it is only a function of remaining time to maturity it is time-homogenous and this specification is capable of producing a hump\(^\text{16}\) shaped instantaneous term structure of volatility which carries over to the term structure of both implied caplet and cap volatilities. The specification therefore fulfils the desired properties.

The parameter d determines the general level of the volatility, and the remaining parameters determine the placement and magnitude of the hump.

The specification allows for an analytical evaluation of the integrated squared volatility, which will prove very useful in the calibration of the standard model.

6.2 Model and empirical implied volatilities

Having chosen a specification of the instantaneous volatility that is capable of reproducing the hump shaped volatility we can capture the shape of the term structure of volatility for a given strike. But as for the Black formula the log normality of forward rates will mean that the model will produce an implied volatility surface that is flat across strikes.

\(^{16}\) And many other plausible shapes
We will now investigate the empirically observed\(^{17}\) volatility surfaces in order to assess this property.

Until around 1994 the volatility surfaces in all currencies were in agreement with the log normal assumption of the Black formula and the standard volatility model by being flat across strikes.

Around 1994, beginning in Japanese Yen, but soon spreading to other currencies, the volatility surface began experiencing non-flat volatilities across strikes.

In this period the volatility was monotonically decreasing, from low to high strikes as illustrated in Figure 3.

**Figure 3:** A typical volatility skew

The monotonically decreasing volatility was observable until the market crisis in the late 1998. During and after this crisis the shape of the non-flat volatility changed into a shape similar to the one illustrated in Figure 4.

\(^{17}\) Joshi & Rebonato (2003)
The volatility was no longer monotonically decreasing; it was smiling, as it had been observed earlier in the equity markets.\textsuperscript{18} It still had a general decreasing shape from low to high strikes, but for the options with the highest strike, the volatility was increasing. Some authors use the term hockey stick shaped for this volatility shape, but we use the term smiling for a non-flat volatility shape which is not monotonically decreasing.

The separation in time of the appearance of the skew and the smile in implied volatilities suggest that two different sources are causing the non-flat volatility. Having established the history of the implied volatility surface of interest rates, we now turn to the current surface.

\textsuperscript{18} Appearing after 1987
We see that the current cap surface\textsuperscript{19} displays both a skew and a smile. The smile is most pronounced for the shortest maturities and is more shallow for the longer maturities.

This leads us to reject the log normality of the standard LIBOR market model, as it prescribes a constant volatility across strikes. This does not imply that we thereby discard the standard model all together. The standard model is intuitive and easy to both calibrate and implement and the usability of the model must be based on a concrete evaluation of the pricing and hedging abilities of the model for each product under consideration.

But the underlying assumption is not fulfilled. The skew and the smile is evidence of a distribution that is skewed and exhibits excess kurtosis.

We will now give a brief overview of different approaches to obtain a non constant volatility across strikes.

\textbf{6.3 Different approaches to obtain a non constant volatility}

Several different methods have been proposed to account for the observed deviations from flat volatility. As the smiling volatility surfaces were first experienced in equity

\textsuperscript{19} Obtained by interpolating the quoted cap matrix
markets, most of these methods have been developed in order to match these smiles, and subsequently been adopted for interest rates. The methods include among others:
- Constant elasticity of variance models
- Displaced diffusion models
- Jump diffusion models
- Regime switching models
- Stochastic volatility models

6.3.1 Constant elasticity of variance models

The constant elasticity of variance (CEV) models take the approach that the volatility of absolute changes in the forward rates is not proportional to the level of the forward rate, as Equation (19) suggests. Still, higher levels of the forward rates lead to a larger volatility, but the scaling is not proportional. Omitting the drift term the CEV model\(^{20}\) writes:

\[
\frac{d f_i(t)}{\sigma_i(t)f_i^\beta} = d z_i(t)
\]

For a geometric Brownian motion the volatility of the increments scale perfectly with the level of the forward rate resulting in a constant volatility of percentage increments, which in turn leads to a log normal distribution.

In the CEV model the parameter \(\beta\) allows a different scaling of the volatility. For \(0<\beta<1\) the volatility of the increments will scale less than proportionally with the level of the forward rate. This means that the volatility of percentage increments will be larger for lower levels of the forward rate than for higher. This leads to a skewed distribution of the logarithm of the forward rate that in turn leads to monotonically decreasing implied volatility as a function of strike.

We see that conditional on the evolution of the forward rate, the volatility is known with certainty, no additional stochastic variables has been introduced. The state of the world at time \(t\) is determined uniquely by the evolution of the forward rate up to time \(t\).

\(^{20}\) Andersen & Andreasen (2000)
6.3.2 Displaced diffusion models

Displaced diffusion models are characterised by instead of modelling the forward rate as a diffusion it models the forward rate plus some constant, $\alpha$, the displacement coefficient. Again omitting the drift term the displaced diffusion reads\textsuperscript{21}

\begin{equation}
\frac{d(f_i(t) + \alpha)}{f_i(t) + \alpha} = \sigma_i(t)dz_i
\end{equation}

We see that the volatility is unconditionally deterministic as in a geometric Brownian motion. But the volatility is the volatility of the percentage increments to the quantity $f_i(t) + \alpha$, not the forward rate itself. For a positive diffusion coefficient the displaced diffusion will produce a monotonically decreasing implied volatility across strikes by altering the scaling of instantaneous volatility with the forward rate, just as the CEV model, and it can be shown\textsuperscript{22} that the CEV and the displaced diffusion model will result in almost identical volatility skews.

The CEV model is preferred over the displaced diffusion theoretically as it ensures positive rates. The displaced diffusion ensures that $f_i(t) + \alpha$ is positive, and the forward rate is therefore only bounded to be larger than $-\alpha$. But the displaced diffusion is preferred over CEV as the CEV model does not provide tractable analytical solutions for caplet prices. In the displaced diffusion model the quantity $f_i(t) + \alpha$ is log normally distributed. This means that all the results linked to the log normality can be retained. This implies i.a. that caplet prices can be calculated from a simple extension of the Black formula\textsuperscript{23}.

Rebonato (2002) shows that for reasonable values of $\alpha$ the probability of experiencing negative rates are rather small, and the resulting difference between a CEV and displaced diffusion model is negligible. Displaced diffusion is therefore preferred for its tractability although CEV is financially more appealing.

\textsuperscript{21} Rubinstein (1983)

\textsuperscript{22} Marris (1999)

\textsuperscript{23} See Appendix 5: Displaced diffusion extension of the Black formula
6.3.3 Jump diffusion models

Jump diffusion models introduce an additional source of pure uncertainty as it introduces a new stochastic variable to the process. A simple example\(^{24}\) could be:

\[
\frac{df_i(t)}{f_i(t)} = \sigma_i(t, T_i)dz_i(t) + d \left( \sum_{i=1}^{N_i} (Y_i - 1) \right)
\]

(31)

where

- \(N_i\) is a Poisson process
- \(Y_i\) is a log normal random variable

The introduction of the jump results in a non log-normal distribution of the forward rates, which produces (dependent on the specific jump-diffusion model chosen) a non-flat volatility across strikes.

6.3.4 Regime switching models

Regime switching models work with two (or more) different processes for the forward rate. Dependent on the state of the world the evolution of the forward rate will be governed by one of the possible processes. A simple example of a regime switching model could be\(^{25}\)

\[
\frac{df_i(t)}{f_i(t)} = \sigma_{i}^{(S_i)}(t, T_i)dz_i(t) \quad \text{for} \quad S = S_i
\]

\[
\frac{df_i(t)}{f_i(t)} = \sigma_{i}^{(S_2)}(t, T_i)dz_i(t) \quad \text{for} \quad S = S_2
\]

(32)

\[P(S = S_1) = p\]
\[P(S = S_2) = 1 - p\]

The model changes, according to the probability \(p\), between two regimes with different volatilities.

Regime switching models can also produce non-flat volatilities.

\(^{24}\) Glasserman & Kou (2000)

\(^{25}\) The authors own formulation
6.3.5 Stochastic volatility models

Stochastic volatility models deviates from the standard LIBOR market model by making the volatility not only a function of time but also a function of additional state variable(s). It is often, as in Heston (1993), assumed that the squared volatility follows a mean reverting diffusion:

\[
\begin{align*}
\frac{df_i(t)}{f_i(t)} &= \sigma_i(t, T_i)dz_{i,1}(t) \\
d\sigma_i^2(t) &= \kappa(\theta - \sigma_i^2(t)) + \nu\sigma_i^2(t)dz_{i,2}(t)
\end{align*}
\]

The stochastic volatility models are capable of producing smiling volatilities as they introduce excess kurtosis in the distribution of the underlying.

Intuitively the origin of the smile can be analysed in the context of a result by Hull & White (1987): When the Brownian motions driving the volatility and the Brownian motions driving the forward rates are uncorrelated the price of a caplet under stochastic volatility is given by the Black price integrated over the volatility distribution:

\[
P_i = \int P^{\text{Black}}_i(T, \sigma^{\text{black}})\phi(\sigma^{\text{black}})d\sigma^{\text{black}}
\]

The caplet price under stochastic volatility will hence be an average of Black prices over the possible values of the Black volatility. ATM the Black formula is approximately linear in \( \sigma^{\text{black}} \). A high volatility state will therefore offset the corresponding low volatility state. Therefore the price according to Equation 34 will not be very different from the price one would obtain using the standard Black formula using the average volatility, leading to an implied volatility of the ATM caplet approximately equal to the average volatility. But as one moves either in or out of the money the Black formula becomes convex in \( \sigma^{\text{black}} \). This means that the average price across volatilities will be higher than the price one would obtain using the average volatility, leading to a higher implied volatility than ATM. Because of the
increasing convexity when moving from ATM to ITM or OTM, stochastic volatility results in a smiling implied volatility.
7 The extended LIBOR market model

In Section 6.2 we saw that the smile in the volatility emerged years after the monotonically decreasing slope. We argued that this could be taken as evidence that we should model the non-flat volatility by two distinct features, one giving rise to a monotonically decreasing volatility and another giving rise to a smile. Joshi & Rebonato (2003) take this approach. Their model is a combination of a displaced diffusion and a stochastic volatility model.

\[
\frac{df_i(t) + \alpha}{f_i(t) + \alpha} = \mu_i^{(a)}(f_i, t)dt + \sigma_i^{(a)}(t)dz_i
\]

\[
\sigma_i^{(a)}(t) = [a_i + b_i(T_i - t)]e^{-r_i(T_i - t)} + d_i
\]

(35)

\[
da_i = RS_a(RL_a - a_i)dt + \sigma_a dz_a
\]
\[
db_i = RS_b(RL_b - b_i)dt + \sigma_b dz_b
\]
\[
d[\ln(c_i)] = RS_c(RL_c - \ln(c_i))dt + \sigma_c dz_c
\]
\[
d[\ln(d_i)] = RS_d(RL_d - \ln(d_i))dt + \sigma_d dz_d
\]

\[
E[dz_i dz_j] = E[dz_i dz_b] = E[dz_i dz_c] = E[dz_i dz_d] = 0
\]

(36)

\[
E[dz_a dz_b] = E[dz_a dz_c] = E[dz_a dz_d] =
\]
\[
E[dz_b dz_c] = E[dz_b dz_d] = E[dz_c dz_d] = 0
\]

(37)

\[
E[dz_i dz_j] = \rho_{i,j}
\]

Superscripts in parenthesis are labels

The displaced diffusion is introduced in the form presented above, modelling the quantity \(f_i(t) + \alpha\) instead of \(f_i(t)\) as a diffusion.

For the stochastic volatility part of the model the route taken is not making the volatility stochastic by adding a random term to the volatility specification itself. Instead it is assuming the same functional form for the instantaneous volatility as in the standard model but the coefficients (or their logarithms) follow Ornstein-Uhlenbeck processes, with reversion level, RL, reversion speed, RS, and volatility \(\sigma\).
As described in section 6.1 the coefficients of the volatility function determine the general level of the volatility and the location and magnitude of the hump. Therefore the stochastic volatility part of the model can be interpreted as allowing the level and location and magnitude of the hump to change stochastically, while preserving the functional form for the volatility.

We notice that the extended model encompasses the standard model. The standard model can be considered a special case with all reversion levels equal to the current level of the coefficients, all volatilities in the Ornstein Uhlenbeck processes set to 0 and \( \alpha = 0 \). Two other interesting special cases exist: The Displaced Diffusion model with no stochastic volatility and the Stochastic Volatility model with no displaced diffusion.\(^{26}\)

When integrating Equation (35) we get an expression for the forward rate at some future point in time very similar to the one presented in Equation (22) for the standard model.

The Ornstein Uhlenbeck process is a mean reverting process. For each increment the drift term will pull the value of the coefficient towards the reversion level, the strength of the pull determined by the reversion speed. Besides this deterministic pull a stochastic term will affect the coefficient, by a Brownian motion scaled by the volatility of the process. The mean reverting behaviour of the process ensures that the coefficients do not deviate “too much” from the reversion level. The concept of time homogeneity is thereby maintained in a slightly different meaning, as we attain that the stochastic volatility fluctuates around the same reversion level in the future as it does today.

The mean reversion causes the average volatility to be more variable for shorter maturities than longer. As a lower variability in the volatility will result in a more shallow smile, the mean reversion will result in the smile being more shallow for longer maturities than shorter.

\(^{26}\) We will return to these special cases later
We thereby see that increasing the volatilities of the coefficients increases the smile, while increasing the reversion speeds makes the smile more shallow for longer maturities than short.

The Ornstein Uhlenbeck process allows an exact closed form solution\(^{27}\) for the value of the coefficient at future points in time. This will prove very useful in the calibration of the model and in the pricing process.

The conditions that the Brownian motions should be pair wise uncorrelated reduces the computational burden, as each coefficient can be evolved independently of the other. Additionally as Brownian motions driving the forward rates are uncorrelated with the Brownian motions driving the volatility the calibration of the model is simplified by the Hull and White result presented in Section 6.3.5.

### 7.1 No arbitrage drifts

As the derivation of the no arbitrage drifts holds for deterministic as well as stochastic volatility the no arbitrage drifts are basically unchanged. Taking into account the displaced diffusion part of the model the no arbitrage drifts, when using \(P(t,T_{j+1})\) as numeraire, becomes:

\[
\begin{align*}
\mu^{(a)}_{i} & = \sigma^{(a)}_{i}(t) \sum_{k=j+1}^{i} \frac{\sigma^{(a)}_{k}(t) \rho_{k}(t)(f_{k}(t) + \alpha) \tau_{k}}{1 + f_{k}(t) \tau_{k}} \quad \text{for } i>j \\
(38) \quad \mu^{(a)}_{i} & = -\sigma^{(a)}_{i}(t) \sum_{k=i+1}^{j} \frac{\sigma^{(a)}_{k}(t) \rho_{k}(t)(f_{k}(t) + \alpha) \tau_{k}}{1 + f_{k}(t) \tau_{k}} \quad \text{for } i<j \\
\mu^{(a)}_{i} & = 0 \quad \text{for } i=j
\end{align*}
\]

We just remember that the volatility is stochastic not deterministic, when integrating the drifts.

\(^{27}\) See Appendix 2: Evolving an Ornstein Uhlenbeck process
7.2 Comparison of the models

We see that in terms of the stochastic differential equations describing the dynamics of the forward rates, the displaced diffusion, stochastic volatility model is an intuitive extension of the standard model. The model adds a skew by modelling \( f_i(t) + \alpha \) instead of \( f_i(t) \) in the diffusion, thereby inducing the same effect as the CEV model, that the volatility of the percentage increments are not constant as in the standard model but decreasing for an increasing forward rate.

The stochastic volatility is introduced by maintaining the parametric form of the volatility, but letting the coefficients of the function be stochastic. This allows an intuitive interpretation of the stochasticity as variations in the general level and the placement and magnitude of the hump of the volatility.

We see that going from the standard to the extended model does not alter significantly the no arbitrage drifts. Although the differences in going from the standard to the extended model at this time seem small, we will now address the issue of implementation, which will reveal some larger differences.
8 Implementation issues

The treatment of the implementation issues will answer the question: “Given the dynamics of the models just presented, how do we actually obtain prices from the models?”

An implementation of the models consists of two parts: Algorithms to calibrate the model, i.e. determining the parameters of the model, and a pricing model. This section will introduce concepts which are central to both parts, but the presentation will be made in the context of the pricing model, while the calibration procedure is discussed in depth in Section 10.

Recalling the pricing formula in Equation (13) we see that the pricing problem amounts to evaluating an expected value under the measure induced by the choice of numeraire.

The pricing procedure thereby comprises of the following steps:
- Choosing the measure/numeraire
- Sample the distribution of the forward rates under the equivalent martingale measure
- Calculating the payoff, taking expected value and transform from relative prices to \( C_t \)

8.1 Choosing the measure/numeraire

We choose to price under the terminal measure of the last caplet in the product. That is we take as numeraire the zero coupon bond maturing at the maturity of the product. This means that the pricing formula considered are Equation (16). The price at time \( t \), \( C(t) \), of a cap maturing at time \( T_n \) can thereby be written as

\[
C(t, T_n) = P(t, T_n)E_{Q}^{F_t} (C(T_n, T_n)|F_t)
\]

We must calculate the expected value of the payoffs at maturity. The expected value must be calculated under the terminal measure. And from the expected value we
obtain the current value of the option by multiplying by the current value of the numeraire, the zero coupon bond maturing at the same time as the cap.

8.2 Sampling the forward rates distribution

For the exotic products we can no longer calculate the expectation analytically as it is possible for standard caps, we must resort to numerical methods. Because of the dimension of the problem the most efficient method is Monte Carlo simulation.

The brute force approach to evolve the forward rates would be to Euler discretize the forward rate and volatility equations and perform a short stepped Monte Carlo simulation of the volatility and the forward rates simultaneously. This would be very computationally inefficient. So we turn to a smarter approach which will reduce the computational burden considerably.

The two products we are interested in pricing, the Ratchet and the Sticky, are characterised by the fact that the payoffs are determined uniquely by the value of the forward rates at their own resets. We can therefore implement the procedure Rebonato (2002) calls the very long jump approach.

The very long jump approach evolves all forward rates to their respective resets in one jump. This affords a considerable reduction in the computation time compared to the brute force approach.

Not all products exhibit this property that the payoffs can be determined solely from the values of the forward rates at their own resets. For instance we considered such an option in Section 2. That options payoff was determined at maturity based on the four year spot rate. In order to determine this payoff we would need the value of the forward rates making up the four year spot rate at the maturity of the option, i.e. before their resets. A model using the very long jump approach would not be able to price this option. Fortunately other efficient implementation techniques exist for these products, which avoid the brute force approach. In short these techniques consists of evolving all forward rates between price sensitive events\footnote{Price sensitive events are points in time where some element concerning the price is determined, e.g. the strike, a payoff etc.} using factor reduction, i.e.
fewer driving Brownian motions than forward rates, and an efficient computation of the drifts which reuses previous computations.\textsuperscript{29}

As we are only interested in pricing the Ratchet and Sticky cap we implement the very long jump procedure, and the following part of the section will show how to implement the approach in order to ensure that the forward rates are evolved according to their joint dynamics.

What we would like to do is to recover a known joint distribution of the forward rates at their resets, as this would yield an easy way to sample the forward rates. The problem is that when we choose one common measure all but one forward rate will have state dependent drifts which results in non lognormal distribution. Additionally for the extended model the volatility is no longer deterministic, which also results in a non lognormal distribution of the forward rates.

Fortunately we can recover the log normal distribution of the forward rates for computational purposes. Concerning the stochastic volatility in the extended model we exploit the result of Hull and White stated in Section 6.3.5: That when the Brownian motions driving the volatility and the Brownian motions driving the forward rates are uncorrelated, as it is in our models, the price of a caplet under stochastic volatility is given by the Black price integrated over the volatility distribution:

\begin{equation}
P_i = \int Black(T_i, V_i) \phi(V_i) dV_i
\end{equation}

In other words because of the independence between the volatility evolution and the forward rate evolution, for each volatility path we can evolve the volatility and treat it as deterministic when evolving the forward rate. This means that given the volatility path the volatility necessary for the evolution of the forward rate can be considered deterministic.

\textsuperscript{29} See Rebonato (2002) and Joshi & Rebonato (2003) for a full description of the method.
As for the drifts we will show later that very accurate approximations exist, in which also the drifts can be considered deterministic conditional on the Brownian motion driving the forward rate.

This means that given a particular realisation of all the Brownian motions in the dynamics the forward rates exhibits deterministic drifts and volatilities, which means that the forward rates are joint log normally distributed at their reset.

These techniques explored further in the following sections imply that the forward rate distributions will be a mixture of log normal distributions with different means and volatilities. The conditional log normality will simplify the sampling of the forward rates significantly.

Having established the joint log normal property of the forward rates we turn to the elements necessary to establish the actual distribution: The covariance matrix and the (deterministic) drifts.

### 8.2.1 The covariance matrix

The covariance matrix is termed the TOtal Terminal Covariance, TOTC, matrix by Rebonato in order to highlight that it is the covariance matrix for the terminal value of the forward rates at their resets.

The TOTC matrix is made up from the elements

\[ M(i, j, k + 1) = \int_{T_k}^{T_{k+1}} \sigma_i(u)\sigma_j(u)\rho_{i,j}(u)du \]

These has the intuitive meaning of being the marginal covariance between forward rate i and j for the time period \( T_k \) to \( T_{k+1} \).

From these elements the TOTC matrix is constructed as:

\[ TOTC(i, j) = \sum_{k=1}^{\min(i,j)} M(i, j, k) \]

The minimum function in the summation means that the sum only runs until either of the forward rates resets.
By writing

\begin{equation}
(TOTC(i,i) = \int_t^T \sigma_i(u) \sigma_j(u) \rho_{ij}(u) du = \int_t^T \sigma_i^2(u) du)
\end{equation}

we note that TOTC(i,i) is the total variance of the i’th forward rate, a result we will use in the calibration of the models.

From the covariance elements of the TOTC matrix we can define the Terminal correlation as the correlation between two forward rates at their respective resets.\(^{30}\)
The terminal correlation is approximately given by

\begin{equation}
\text{Terminal correlation}_{i,j}(t,T) \approx \frac{\int_t^T \sigma_i(u) \sigma_j(u) \rho_{ij}(u) du}{\sqrt{\int_t^T \sigma_i^2(u) du \int_0^T \sigma_j^2(u) du}}
\end{equation}

Terminal correlation describes the correlation between the forward rates at their individual resets. As we integrate the product of instantaneous volatilities and correlations we see that the terminal correlation between forward rates is determined not only by the instantaneous correlation but also the “distribution” through time of the instantaneous volatilities. This means that the volatility function chosen and the parameters obtained by the calibration will affect the correlation between forward rates at their resets.

Returning to the TOTC we will now present procedures to actually calculate the covariance integrals.

For the standard model where the instantaneous volatility function is given by Equation (28), an analytical solution to the integrals exists.\(^{31}\) The solution is provided in Appendix 3: The volatility integral in the standard model.

\(^{30}\) Brigo & Mercurio (2006)

\(^{31}\) Given a constant instantaneous correlation, as we work with
For the Extended model with stochastic volatility there is no analytical solution and we must resort to numerical methods.

The integrals are sampled numerically by dividing the integration period into a number, \( v \), of sufficiently small steps \( \Delta s = \frac{T_n - t}{v} \). The volatility is then evolved by evolving the volatility coefficients, \( a, b, c \) and \( d \) over the time step. As the coefficients (or their logs) are driven by Ornstein Uhlenbeck processes they can be evolved exactly over any time step.\(^{32}\) So no discretization error is introduced in this step.

Given the evolution of the coefficients we can calculate for all the forward rates the value of the instantaneous volatility function corresponding to their remaining time to maturity.\(^{33}\) This enables us to calculate the marginal covariances and from them the total terminal covariances:

\[
TOTC(i, j) = \sum_{l=1}^{\min\left\{ \frac{T_i - T_j}{\Delta s}, \frac{T_j - T_i}{\Delta s} \right\}} \sigma_i ((l - 0.5) \Delta s) \sigma_j ((l - 0.5) \Delta s) \rho_{ij} \Delta s
\]

Though the expression is quite messy, the interpretation is straightforward. The sum runs from the first time step to what corresponds to the upper bound of the integral, and the instantaneous volatilities are evaluated at the midpoint of the time steps.

Although the coefficients can be evolved exactly over any time step, the same does not go for the volatility integral, the procedure will introduce a discretization error. This discretization error will go to zero as we decrease the size of the time step in the Monte Carlo simulation, but this will increase the computation time. The discretization error will be examined in Section 11.

We see that the difference between the standard and the extended model lies in how to obtain TOTC. For the standard model the volatility is deterministic and we can calculate TOTC analytically using closed form solutions. And as the volatility is deterministic we only need to calculate it once.

\(^{32}\) See Appendix 2: Evolving an Ornstein Uhlenbeck process

\(^{33}\) We calculate the instantaneous volatility for a time to maturity matching the midpoint of the time step we integrate over.
For the extended model we need to calculate TOTC for a number of volatility paths because of the stochastic nature of the volatility. Additionally as the coefficients vary stochastically over the volatility path TOTC must be calculated by numerical methods.

This means that the extended model is much more computationally demanding than the standard model. The TOTC matrix is used in the calibration of the models, where we minimize a penalty function using an iterative search. For the standard model this is not a problem because of a fast computation of TOTC due to the closed form solution. But for the extended model the repeated numerical sampling of a possibly large number of TOTC realisations means that the calibration procedure is much more computationally demanding compared to the standard model.

### 8.2.1.1 Low discrepancy numbers

As the use of numerical methods in the calibration comes at a high computational cost, we seek to reduce the number of volatility paths necessary to obtain the required degree of accuracy.

In ordinary Monte Carlo simulation, we draw (pseudo) random numbers uniformly distributed between 0 and 1 that we translate to the distribution of which we want to calculate some integral. The problem is that we introduce numerical error if the random numbers we draw are not representative of the uniform distribution as it will transform into a distribution not representative of the problem. For example we could by chance draw too many numbers above 0,5, resulting in a biased distribution. By the central limiting theorem we know that as we increase the number of draws/paths the numerical integral will converge towards the true value. But instead of increasing the number of pseudo random numbers to obtain an even distribution in the interval, we turn to a set of numbers that are constructed in order to be as evenly distributed as possible in the 0;1 interval.

Low discrepancy numbers (LDN) are series of numbers with this property. For a given number of draws the low discrepancy will by construction be evenly distributed over the interval, whereas we cannot be certain that this is the case with pseudo random numbers (Glasserman (2004)).

The principle of a LDN sequence is that the next number in the sequence is located as far away as possible from the preceding numbers, thereby reducing clustering of the
numbers. The sequence thereby makes the next number add as much information to
the evaluation of the problem as possible as it does not replicate any information from
the previous numbers. To illustrate the concept, Table 1 gives the first ten numbers of
the van der Corput (vdC) sequence in base 2.

Table 1: Van der Corput sequenze in base 2

<table>
<thead>
<tr>
<th>First 10 numbers of the van der Corput sequence in base 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,5</td>
</tr>
<tr>
<td>0,25</td>
</tr>
<tr>
<td>0,75</td>
</tr>
<tr>
<td>0,125</td>
</tr>
<tr>
<td>0,625</td>
</tr>
<tr>
<td>0,375</td>
</tr>
<tr>
<td>0,875</td>
</tr>
<tr>
<td>0,0625</td>
</tr>
<tr>
<td>0,5625</td>
</tr>
<tr>
<td>0,3125</td>
</tr>
</tbody>
</table>

We see that the numbers are evenly distributed over the 0;1 interval.

The base of the vdC sequence can be any prime and two vdC sequences in different
bases will be uncorrelated, so that the two sequences are distributed evenly over a
square. When combining multiple vdC sequences they make up a Halton sequence in
n dimensions.

For the volatility evolution of our model we need four sets of random numbers, one
for each volatility process. These four sets of random numbers should be uncorrelated
both with each other and across time, while being evenly distributed across paths.
This means that we would need four times the number of time steps different
dimensions. As only the first 6-8 vdC sequences of a the Halton sequence are usable
we cannot use a new vdC sequence for each time step.

Other sequences\(^{34}\) exist but without going to commercial (read: expensive) versions\(^{35}\)
the dimensionality problem cannot be overcome by choosing another sequence.

---

34 One of the mostly used is the Sobol sequence
35 A 1024-dimensional sequence without high dimensional problems is available at www.broda.co.uk
Instead we employ the hybrid quasi-Monte Carlo method as suggested by Glasserman (2004). The implementation follows Dias. The method uses one vdC sequence in a different dimension for each coefficient to ensure that the numbers driving each coefficient are uncorrelated. The sequence constitutes the first time step, as it evolves the coefficient over the first time step for each path, thereby ensuring that the paths are evenly distributed. To get the next time steps, we need sequences that are uncorrelated with the first. This is obtained by doing random permutations of the first sequence. By doing this the numbers are uncorrelated along the path and evenly distributed across paths, fulfilling Equation (36) while allowing fast convergence.

The vdC sequences are uniformly distributed from 0 to 1. In order to evolve the coefficients in accordance with the Ornstein Uhlenbeck dynamics, we need normally distributed sequences. This could be done by Excel's built-in function to invert the cumulative normal distribution, Normsinv(). But as this function has been shown to have a really bad performance in the tails and is rather slow, we use Moro’s inversion.36

Having determined the covariance matrix of the joint distribution we now turn to the drifts.

8.2.2 Conditionally deterministic drift terms

What we want to do is to construct an approximation that removes the state dependency from the drift terms as this will mean a log normal distribution of the forward rates.

The state dependence stems from the forward rates and, for the extended model, the volatilities appearing in the drifts. As stated above, after we have sampled the volatility we can regard it as deterministic with regard to the forward rate evolution. The issue of state dependence in the drifts therefore only concerns the forward rates.

A (naive) first try would be the well known Euler approximation, where we set the forward rates in the drift equal to their current values during the entire time step.

36 See Moro (1995)
This can be written as:

\[
\int_t^T \left[ \sigma_i(u) \sum_{k=j+1}^i \frac{\sigma_k(u) \rho_{ik}(u) f_k(u) \tau_k}{1 + f_k(u) \tau_k} \right] du \\
= \int_t^T \left[ \sigma_i(u) \sum_{k=j+1}^i \frac{\sigma_k(u) \rho_{ik}(u) f_k(0) \tau_k}{1 + f_k(0) \tau_k} \right] du \\
= \sum_{k=j+1}^i \frac{f_k(0) \tau_k}{1 + f_k(0) \tau_k} \int_t^T [\sigma_i(u) \sigma_k(u) \rho_{ik}(u)] du
\]  

(46)

The last rewrite reveals that the drifts can reuse information from the volatility calculations as the integral is just TOTC(i,j).

Rebonato (2002) states that this approximation is fairly accurate for a step size of up to a year. This means that the approximation is not acceptable for the very long jump procedure.

The approximation applied in this implementation is the Predictor-Corrector approximation proposed by Hunter, Jäckel & Joshi (2001). For simplicity we first show a general case of the Predictor-Corrector method, followed by an application of the method to our problem.

We assume a stochastic differential equation for the process \(Y(t)\) given by

\[
dY(t) = \mu(Y(t), t) dt + \sigma(t) dz(t)
\]  

(47)

We see that the drifts are state dependent and the volatility is not. The Euler approximation is then given by:

\[
Y(t + \Delta t) \equiv Y(t) + \mu(Y(t), t) \Delta t + \sigma(t) \Delta z(t)
\]  

(48)

37 Following the presentation in Rebonato (2002)

38 As is the case in our problem given a particular volatility path
The Predictor Corrector approximation states that the process can be approximated by

\[
Y(t + \Delta t) \approx Y(t) + \frac{1}{2} (\mu(Y(t), t + \Delta t) + \mu(\hat{Y}(t + \Delta t), t + \Delta t) + \sigma(t) \Delta z(t))
\]

(49)

where

\[
\hat{Y}(t + \Delta t) \approx Y(t) + \mu(Y(t), t) \Delta t + \sigma(t) \Delta z(t)
\]

This approximation first evolves the process using the Euler approximation to obtain a first estimate of \(Y(t + \Delta t), \hat{Y}(t + \Delta t)\). This is then used to calculate \(\mu(\hat{Y}(t), t)\), the drift corresponding to the first estimate of the forward rate at the end of the step. The final value of the forward rate is obtained by evolving the forward rate using the average drift, approximated by \(\frac{1}{2} (\mu(Y(t), t + \mu(\hat{Y}(t + \Delta t), t))\). It should be noted that \(\Delta z(t)\) is the same in Equation (48) and Equation (49) to ensure that the first and final path experience the same shocks.

When applied to our model the method is implemented by first evolving all forward using the Euler approximation in Equation (46), to obtain first estimates of the forward rates at their resets, \(\hat{f}_i\). These forward rates are used to calculate new drifts again according to the deterministic approximation in Equation (46). The final values of the forward rates at their resets are then obtained using the average of the two drifts.

The advantage of the Predictor Corrector approximation is that it recovers the log normal property of the forward rates. Conditional on the increments to the Brownian motions the drifts are deterministic. Having drawn a particular set of random numbers to generate the Brownian increments the value of the drifts can be calculated. Both the first Euler and the following Predictor Corrector step have a deterministic drift and the forward rates can easily be obtained based on their log normal properties. The resulting drifts are consequently called conditionally deterministic; conditional on the Brownian increments the drifts are deterministic. And thereby we recover the log normal property of the forward rates for computational purposes. As stated the before
the unconditional distribution of the forward rates will be a mixture of log normal distributions with different means.

Three approximations are embedded in the Predictor-Corrector approximation. First we hope that the first estimate, $\hat{f}_i$, will be near the actual value of the forward rate at the end of the time step. If this is not the case the averaging will not improve the drift considerably. Second we hope that the average drift can be approximated by

$$\frac{1}{2} \left[ \mu(Y(t), t) + \mu(\tilde{Y}(t + \Delta t), t) \right].$$

By using this approximation we ignore what happens to the drift during the time step. Finally we hope that substituting the state-dependent drift with a deterministic one will not distort the dynamics of the forward rate. Rebonato (2002) shows that the Predictor-Corrector approximation affords a very accurate approximation, which allows us to evolve the forward rates over time steps as long as 20 years. And the method is therefore applicable for the very long jump approach.

As we have covered a lot of ground concerning the implementation, we will now restate the steps in the sampling of the forward rate realizations. We base the presentation on the expression for a forward rate at some future time, $T$, which in the case of the extended model can be written in terms of the TOTC elements and using Cholesky decomposition\(^3\) as

$$f_i(T) = (f_i(t) + \alpha) \exp \left( \mu_i^{(PC)} + TOTC(i, i) + \sum_{k=1}^{n} A_{i,k} Z_k \right) - \alpha$$

(50)

where

- $\mu_i^{(PC)}$ is the drift integral obtained from the Predictor Corrector method
- $AA^T = TOTC$
- $Z$ is a vector of uncorrelated normal random variables

\(^3\) See Appendix 4: Cholesky decomposition
We first build the TOTC matrix. We start by evolving the volatility coefficients using low discrepancy numbers. From the evolution of the volatility coefficients we can evaluate the covariance integrals necessary for the TOTC. We then do a Cholesky decomposition of the TOTC matrix, which combined with a vector of uncorrelated Brownian motions gives us the diffusion parts of the equations, as normally distributed shock with a covariance according to TOTC.

To obtain the drift term we employ the Predictor Corrector method. Keeping the diffusion part fixed, we first evolve the forward rates to their resets by the Euler approximation remembering that the covariance integrals in the drift calculation can be obtained directly from the TOTC matrix. We then recalculate the drifts using the first estimates of the forward rates at their reset. The average between this drift and the Euler drift is then used in the final evolution of the forward rates again reusing the diffusion part.

For the standard model only one realisation of the TOTC matrix is possible, and we therefore only need to sample one volatility path.

In the extended model different realisations of TOTC is possible and we must therefore sample a number of different volatility paths. For each volatility realisation i.e. TOTC matrix, we calculate a number of different forward rate paths in order to reduce the overhead in establishing the TOTC matrix. This means that instead of sampling a new TOTC matrix for each forward rate path we sample multiple forward rate paths given the same TOTC matrix. This procedure follows the idea from Equation (33), but with a numerical integration over forward rate paths.

### 8.3 Payoff and expected value

For each forward rate path sampled we have then obtained all the necessary forward rates at their reset. This makes it straightforward to calculate both the strikes and payoffs of the products using the payoff and strike definitions of Equation (6), (9) and (10).
Due to the concept of time value of money the payoffs can not just be summed as they are received at different points in time. As the expectation is taken of the time $T_n$ value of the payoffs, we must roll up all payoffs to that time.

We know that when we receive a payoff at time $T_i$, the payoff can be invested at the rate of the forward rates as they reset along the path. The payoff received at time $T_i$ can thereby be rolled up to time $T_n$ by the following formula:

$$c(T_n, T_i) = c(T_i, T_i) \prod_{k=i}^{n-1} (1 + f_k \tau)$$

(51) where

$c(T_n, T_i)$ is the time $T_n$ value of the payoff received at time $T_i$

$f_k$ are the realisations of the forward rates along the path

The intuition of the formula is clear when remembering that the forward rates are indexed by the reset time.

As the payoffs are now made up at the same point in time we can sum the individual payoffs and estimate the expected value as the average over the paths.

Finally we go from the relative value to the cap value by multiplying by the current value of the numeraire.

We have now established the theoretical framework and addressed a number of issues concerning the implementation of the models.

We will now turn to the empirical part of the paper, where we calibrate the models to market data and analyze pricing differences between models.
9 Data

For the calibration of the models we need forward rates, volatilities and correlations, but none of the desired quantities are quoted directly in the market; we need them from other quantities.
The models are calibrated to the EUR interest rates of 12 September 2006, as obtained from Reuters.\textsuperscript{40}

The current forward rates of 12 September are obtained from a set of swap rates with maturities from 1 to 30 years. The short maturity swaps are not particularly liquid, and more liquid instruments\textsuperscript{41} could produce a more accurate yield curve by reducing the problems of using quotes of illiquid instruments. But using different instruments would introduce problems concerning the comparability of different instruments, i.e. the problem that despite of theoretic arbitrage relations, the quotes are not completely comparable because of e.g. supply and demand issues for the particular instruments. The estimation of the yield curve will therefore be based exclusively on swap rates, but we will keep in mind that we could thereby introduce some noise in the yield curve and consequently in the cap prices.
We use the midpoint between the quoted bid and ask swap rate.

The volatility is calibrated to the volatility matrix of EUR caps. The cap implied volatility quotes consists of 16 maturities ranging from 1 to 30 years,\textsuperscript{42} and 14 strikes from 1,5% to 10,0%.
Though we can obtain quotes for the strike range above, the caps far ITM or OTM must be considered very illiquid and are products of an extrapolation method rather than reliable prices. We therefore reduce the calibration sample to a subset of the strikes. With a rather flat yield curve between 4,0 and 4,5 %, we assess that strikes from 3,0% to 7,0% can be considered liquid enough for calibration purposes.

\textsuperscript{40} Thanks to Scanrate Financial Systems and Jyske Bank for providing the data
\textsuperscript{41} Like FRAs and futures
\textsuperscript{42} The 1Y and 1,5Y caps are with 3 month tenor, but we will use these volatilities as proxies for the 6 month tenor 1Y and 1,5Y caps
Also as the longest maturity considered in the comparison of prices is 10 years, we
assess that the volatility information needed can be extracted from the maturities
between 1 and 20 years.

The volatilities quoted in the market are implied volatilities for the caps,
$\sigma_{\text{Black (Cap)}}(T)$, i.e. the one volatility that put into the Black formula for all the
underlying caplets will result in the market price of the cap. But in order to calibrate
the model we need the implied volatilities of caplets, $\sigma_{\text{Black (Caplet)}}(T)$. We therefore
first describe how to obtain the caplet volatilities from cap volatilities.

The cap implied volatilities will be some average of the underlying caplet implied
volatilities. We therefore need some algorithm to extract the caplet implied volatilities
from the quoted cap implied volatility.

If we had cap implied volatility quotes for all 59 possible maturities from 1 to 30
years, we could bootstrap the caplet implied volatilities from these quotes, as
described later. But as only 16 are quoted we first need to fill in the remaining implied
volatilities by interpolation.

The behaviour of the cap implied volatilities between two interpolation points are
fundamentally determined by the model governing the caplet prices, so the correct
way to obtain the remaining cap implied volatilities would be from an iterative
procedure: Starting from some guess of the interpolation scheme, we would calculate
the cap volatility matrix, the model could then be calibrated to this volatility matrix,
and a new cap volatility matrix could be obtained from interpolation in accordance to
the calibrated model. By continuing the iteration until agreement has been reached
between the interpolated and the model volatility matrix one would obtain an
interpolation of the cap volatility matrix that is model consistent. This would be a very
cumbersome way to obtain the volatility matrix, and we therefore use a simple
interpolation scheme to obtain the full matrix of cap implied volatilities. We use linear
interpolation for its simplicity, but we will keep in mind that a more advanced
interpolation technique could reduce some of the distortion that we will see when
calibrating the model. The interpolated cap surface is showed in Section 6.2.
From the interpolated cap implied volatility matrix we can bootstrap the caplet implied volatility matrix. This is done by recalling that a cap implied volatility is some average, yet to be defined, of the underlying caplet implied volatilities. The first 1 year cap consists of one caplet and the 1 year caplet implied volatility is equal to the 1 year cap implied volatility. For the 1,5 year caplet the implied volatility can be obtained from

\[
(52) \quad \sigma_{Black\ (Cap)}(1,5) = \text{avg} \left( \sigma_{Black\ (Caplet)}(1); \sigma_{Black\ (Caplet)}(1,5) \right)
\]

The 1,5 year caplet implied volatility is the only unknown, and it can therefore be solved for. This bootstrapping procedure can be repeated for each cap to obtain the full caplet implied volatility matrix.

The averaging that produces the cap implied volatility from the caplet implied volatility is proved by Alexander (2002) to be approximately the vega\(^{\text{43}}\) weighted average of the caplet implied volatilities. As vega is a function of the caplet implied volatility, this procedure would again require an iterative procedure, though not as demanding as the full model iteration. To avoid the iterative procedure we assume equal vegas of all caplets, thereby reducing the average to the simple average of the caplet implied volatilities:

\[
(53) \quad \sigma_{Black\ (Cap)}(1,5) = \frac{\sigma_{Black\ (Caplet)}(1) + \sigma_{Black\ (Caplet)}(1,5)}{2}
\]

Alexander shows that while biased, the procedure produces a fairly accurate bootstrapping.

The bootstrap provides us with a full caplet volatility matrix with maturities from 1 to 30 and strikes from 3.0 to 7.0%. The corresponding volatility surface has the following shape.

---

\(^{43}\) Vega is the derivative of the Black premium with respect to volatility
Some observations can be made from the graph. We see a remarkable jagged shape of the volatility surface. This shape is a result of the linear interpolation procedure for the cap volatility matrix, and shows that the linear interpolation is too simple. A practical implementation of the model would probably require a more careful procedure to obtain the caplet volatilities. But we maintain the procedure for simplicity and address the issues resulting from the jagged form when they are encountered.

We see that this subsample of caplet implied volatilities exhibits the same general shape as the full cap surface describes earlier: It exhibits a skew and a smile, and the smile is most pronounced for the short maturities.

The correlation of the models is calibrated to a historic correlation matrix. Instead of using the historic correlation matrix, which is backwards looking and therefore not per se equivalent to the forward looking correlation needed for pricing, it would be...
possible to extract the correlation matrix from swaption quotes. Swaptions are options that give the holder the right to enter a swap with a predetermined fixed rate. As such they depend on the correlation between forward rates, and if the correlation implied by the swaption quotes could be extracted we would have the market view on the future correlations.

Unfortunately the swaption dependence on the correlations is rather limited, which means that it is difficult to extract the correlation information from the swaptions. We therefore calibrate the correlation of the models to a historic correlation matrix.

The historic correlation matrix is obtained from a time series of swap quotes consisting of maturities ranging from 2 to 30 years for the period 01-06-2005 to 11-09-2006.
10 Calibration

The term calibration covers the procedure of determining the parameters of the models given the data described in the previous section. Although strictly speaking the term calibration refers to procedures when the parameters are obtained from quoted market priced at a given point in time, this section will also contain the parameters obtained by estimation, that is parameters obtained from a time series of observations.

We will in turn consider the three inputs of the model; the yield curve, the correlations and the volatilities. Each subsection will develop the calibration/estimation procedure and present the results.

10.1 The yield curve

The LIBOR market model takes as input the current yield curve, through both the discount function \( P(t,T_i) \) and the forward rates \( f(t,T_i) \).

When estimating the yield curve from market quotes there are two ways to go. One way is to bootstrap the yield curve and thereby obtain an exact fit between the prices from the model yield curve and the market quotes. In order to do so some interpolation technique must be employed as more forward rates are needed than quotes are available. This procedure often results in very ill-behaved yield curves and especially the shape of the forward curve will be very difficult to justify financially.

Another way is to impose some parametric structure on the yield curve. By imposing a certain structure it is possible to control the range of shapes the yield curve can take, and thereby insure that the curve(s) will be well behaved. But the cost of this is that the prices from a parametric yield curve model do not necessarily match the market quotes exactly.

At a first glance the inability of a parametric yield curve model to match the observed market quotes looks rather incriminating, but market frictions can justify smoothing the yield curve. The term non-synchronous (or discrete) trading refers to the fact that the trades in the products of a market, does not happen simultaneously. Rather there are some time lag between a trade in one product and a trade in another. The same carries over to the quotes in a market-maker driven market, where not all quotes are updated simultaneously. This means that the market quotes at any given point in time
do not correspond to the same point in time; there are stale quotes. To reduce the
effects of the stale quotes (which are not identifiable) on the yield curve, smoothing
by a parametric yield curve model is financially justifiable.

10.1.1 Nelson-Siegel

To obtain a yield curve that is well behaved for both the discount function and the
forward rate we use the Nelson-Siegel model\(^{44}\) of the spot rate with annual
compounding:

\[
R(t, T) = \beta_0 + (\beta_1 + \beta_2) \times \left(1 - e^{-T/\tau}\right) - \beta_2 \times e^{-T/\tau}
\]

The model gives the spot rate \( R(t, T) \) as a function of time to maturity (T). \( \beta_0, \beta_1, \beta_2 \)
and \( \tau \) are coefficients to be estimated.

The model allows all the shapes that yield curves normally take: Monotonic, humped
and S-shaped.

From the spot rate given by the model it is straightforward to obtain the discount
factors and forward rates as needed using Equation (1) and Equation (2).

10.1.2 Estimation

To estimate the parameters of the model we minimize over the parameters the squared
differences between the market and model swap rates.

\[
\text{Penalty function} = \sum (R^\text{Market}_S - R^\text{Model}_S)^2
\]

This results in the following parameters

| Table 2: Nelson-Siegel yield curve parameters |
|---|---|---|---|
| tau | Beta0 | Beta1 | Beta2 |
| 74,146 | -13,843 | 17,632 | 24,219 |

\(^{44}\) Nelson & Siegel (1987)
This provides us with the following forward rate curve

**Figure 7:** Forward rate curve 12. September 2006

We see that the yield curve is relatively flat around 4,0-4,5%.

Comparing model and market swap rates we see a quite satisfying fit when keeping in mind that the Nelson Siegel parameterisation is rather simple.

**Figure 8:** Market and model swap rates
Apart from a few maturities all swap rates are within the bid-ask spread of 3 bp. But we see some systematism in the errors. This systematism could introduce some bias and if the model is implemented for a trading purpose requiring sharp prices it should be analyzed. But for the purpose of the analysis in this paper, where we will consider general pricing differences, the calibration errors are not a problem.

10.2 Correlation

As mentioned earlier the correlation of the models is obtained from a historic correlation matrix. The historic correlation matrix is obtained as described by Gatarek, Bachert & Maksymiuik (2006) by first estimating the daily yield curves for each observation in the time series of swap rates using the procedure just described. From these yield curves, the daily changes of the logarithm of each forward rate are calculated. Finally the elements of the correlation matrix are calculated as the correlation coefficient for the changes of the pairs of forward rates. The correlation coefficients are calculated for forward rate resets ranging from 0.5 to 30 years. This historic correlation matrix could be utilized directly, but as obtaining a reliable estimate of the correlation matrix is extremely difficult, smoothing of the matrix by using a parametric function is even more warranted when calibrating the correlation than in the case of the yield curve.
One of the simplest parametric forms is proposed by Rebonato (2002) as

\[ \rho_{ij} = \exp(-\beta \cdot |T_i - T_j|) \]

(56)

where

- \( \beta \) is a constant to be estimated
- \( T_i \) is the maturity time of the \( i \)th forward rate

For this parameterization the instantaneous correlation is a function of only the time between the maturity dates of the two forward rates. This parametric function thereby imply that the instantaneous correlation between the forward rates maturing in 1,5 and 2 years is the same as between the forward rates maturing in 29,5 and 30 years. One could argue that this implication is too coarse, that one would expect longer maturity forward rates to experience a higher correlation, than shorter maturity forward rates. This has led to (among others) the parametric specification by Schoenmakers & Coffey (2003). The specification is not reproduced here as the function is very tedious and non intuitive. But the implications follow the above as it allows for a correlation structure which exhibits a stronger correlation between longer maturity forward rates than the shorter maturities.

In spite of the discussion above we will proceed with the parameterization in Equation (56). Although the Schoenmakers and Coffey specification affords a better fit to the estimated correlation, the ease of implementation and the intuitive function makes Rebonatos specification better suited for the analysis in this paper.45

In practice the smoothing is done by minimizing the squared differences between the model correlation matrix and the historic correlation matrix over the parameter \( \beta \).

\[
\text{Penalty function} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(\rho_{i,j}^{\text{Historic}} - \rho_{i,j}^{\text{Model}})^2}{n^2}
\]

(57)

where

- \( n \) is the number of forward rates

45 Other parameterizations proposed by Rebonato (2002) has been examined, but the same goes for these specifications

64
The calibration results in the following correlation parameter

<table>
<thead>
<tr>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,020</td>
</tr>
</tbody>
</table>

And the following correlation surface

**Figure 10:** Historically estimated correlation surface after smoothing

10.3 **Volatility**

As the calibration procedure is very different for the standard and the extended model, we will treat the two calibration procedures separately.

10.3.1 **Calibration of the standard model**

In Section 5 we saw that if we choose the volatility so that

\[ \sigma^2_{\text{Black}}(T)T = \int_t^T \sigma^2(u)du \]
we will recover exactly the quoted prices.
But we have assumed a parametric function for the instantaneous volatility and therefore we cannot be certain to obtain equality for all caplets.
Actually it would be possible to perform a non-parametric calibration. This could be done by bootstrapping the caplet volatilities according to the condition above, thereby recovering the market prices by construction. But the arguments for a parametric form for the volatilities follow the same lines as for the yield curve, and we therefore maintain the parameterization of the instantaneous volatility.
The calibration therefore consists of two steps: First we obtain the best fit parameters of the functional form and second we fine tune to obtain an exact fit.

As the volatility across strikes is constant in the standard model, one often restricts the calibration instruments to contain only one caplet for each maturity. This way the volatility is fitted to a certain strike instead of some average across strikes. The choice of strike is often determined by the character of the product that one subsequently wishes to price. For a product that resembles and is most naturally hedged by ATM caps it would be natural to choose as calibration instruments the ATM caplets.
We will use as calibration instruments the caplets with a cap rate of 4,0%.

In the first step of the calibration we search for the parameterization of the volatility function that provides us with the best fit to the market volatilities.
We therefore define the following penalty function for the standard model

$$\text{Penalty function}_{\text{Standard model}} = \frac{\sum_{i=1}^{n} \left( \int_{t_i}^{T} \sigma(u)^2 \, du - \sigma_{\text{Black}}^2(T_i)T_i \right)^2}{n}$$

(59)

where
n is the number of caplets

The penalty function is minimized over the parameters a, b, c and d, no restrictions are imposed on the parameters. As we have seen the integral is equal to TOTC(i,i) and
it can be calculated using a closed form solution. This means that the minimization can be done in a few seconds.

The calibration results in the following parameters of the instantaneous volatility function:

Table 4: Volatility parameters
(Standard model)

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-6.771</td>
<td>25.046</td>
<td>0.735</td>
<td>10.204</td>
</tr>
</tbody>
</table>

This gives us the following term structure of volatilities

Figure 11: Market and model implied volatility
(Standard model)

We see that the general fit is very convincing, but as expected the parameterized volatility function does not recover exactly the quoted market prices. We therefore

46 See Appendix 3: The volatility integral in the standard model
47 For a graph of the instantaneous volatility function see the spreadsheet Volatility.xls
move to the second part of the calibration, obtaining exact recovery of the quoted prices.

We obtain the exact fit by introducing forward rate specific factors, $\varepsilon_i$.

By setting

$$\varepsilon_i^2 = \frac{\sigma_{\text{black}}^2(T_i)T_i}{\int_s^T \sigma(u)^2 \, du}$$

and letting the volatility of the model be given by

$$\sigma_i^{\text{forward rate specific}} = \varepsilon_i \sigma_i$$

we obtain identity between model and quoted prices.

The consequence of the forward rate specific factors is that the model assumes a constant misalignment between the parametric and true volatility, i.e. if the parametric volatility of the 5 year forward rate is 5% below the true/market implied volatility, we will increase today the volatility of the 5 year forward rate by 5% and as we evolve the forward rate through time in 4 years time the volatility of the 1 year forward rate will also be increased by 5% as compared to the parametric volatility. If the forward rate specific factors are not equal across maturities this implies that the volatility is no longer time homogenous, we cannot be sure that the 1 year forward rate volatility today is equal to the volatility 4 years from now of the forward rate maturing in 5 years.
The forward rate specific terms reveal that the parameterization of the volatility generally allows a satisfying fit as all but two of the forward rate specific terms are between 0.95 and 1.05. The forward rate specific terms correspond to errors in implied volatility of maximum 0.72, when removing the two largest errors. This is quite satisfying when comparing to the results of Rebonato (2002). Unfortunately we also see that there is some systematism in the forward rate specific terms, which indicate that some characteristics of the market implied volatilities are not captured by the model. Fortunately this systematism stems from the linear interpolation scheme, so it is not some underlying financial dynamics that the model is incapable of capturing. This observation leads us to conclude that the linear interpolation scheme employed would be too simple in a practical implementation of the model. We will maintain the linear interpolation scheme, but proceed without the forward rate specific terms. We believe that these will just introduce noise from the interpolation. This approach is valid in the scope of this thesis as we are interested in comparing the general dynamics of the models. We therefore favour a smooth, financial plausibility term structure over the exact recovery of the calibration instruments. Additionally by omitting the forward rate specific terms we ensure a time homogenous volatility.
Though in a practical implementation of the model a more satisfying interpolation scheme (e.g. a splining technique) must be applied, which would render the forward rate specific terms useful.

10.3.2 Calibration of the extended model

When going to the extended model the link between the model instantaneous volatility and the Black implied volatility present in the standard model is broken. When we assume the forward rate dynamics in the extended model, the Black implied volatilities has no financial interpretation except that it is the number that we must put in the Black formula to obtain the correct price.

Still the penalty function will be based on implied volatilities, but only as representations of the option prices, not with any financial interpretation. The procedure is therefore to calculate the model premiums given the volatility parameters, convert these premiums to implied volatilities using the Black formula, and compare these implied volatilities to the implied volatilities of the market prices. It would be possible to abandon the implied volatility concept and instead compare the model premiums to the ones quoted in the market. But the advantage of working with implied volatilities is that these do not exhibit the same variability as the premiums. For a flat implied volatility surface, we would see a great difference in the premiums of a 1 year and a 20 year cap. These differences are reduced by considering implied volatilities instead.\(^{48}\) The penalty function to be minimized is therefore

\[
Penaltyfunction_{\text{Extended model}} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \sigma_{\text{Black}}^{(\text{Market})}(T_i, R_{K,j}) - \sigma_{\text{Black}}^{(\text{Model})}(T_i, R_{K,j}) \right)^2
\]

(62) where

\(m\) is the number of quoted strikes
\(n\) is the number of caplet maturities

Since the extended model can utilize information across strikes, we now calibrate to the full implied volatility matrix.

\(^{48}\) A penalty function based on premiums was explored without success
As in the case of the standard model we generally do not obtain a perfect fit by this calibration. For the extended model we maintain the possibility of introducing forward rate specific terms, which could be chosen either to fit exactly the volatilities of a certain strike or to reduce the overall pricing errors across strikes. As we have previously decided to perform the analysis without the forward rate specific terms in order to avoid the introduction of noise from the interpolation and maintain the time homogeneity, we will not explore the procedure for the extended model, which follows the same lines as for the standard model.

In order to calculate the premiums from which we calculate the model implied volatilities we follow the lines presented in Section 8, but avoid the numerical sampling of the forward rates.

We start by sampling a volatility path, which can then be considered deterministic, when calculating the caplet price. We then change the measure for each forward rate to its terminal measure making them driftless and as the volatility is considered deterministic the forward rates conditional on the volatility path are lognormal and the caplet value is given by the Black formula with Black volatility equal to the root mean squared instantaneous volatility. We consequently need to sample only the volatility numerically, while the caplet payoffs are integrated analytically.

By sampling a number of volatility paths we obtain the caplet value under stochastic volatility as an average of Black prices.

In terms of the TOTC matrix the Black volatility is given by

\[
\sigma_{\text{Black}}(T_i, R_{K,i}) = \sqrt{\frac{\text{TOTC}(i,i)}{T_i}}
\]

This shows that we need only compute the diagonal elements of TOTC matrix.

As we assume the displaced diffusion we must use the displaced diffusion extension of the Black formula given in Appendix 5: Displaced diffusion extension of the Black formula.

In order to reduce the computation time, we reuse as much information as possible. For each volatility path we calculate all maturities using different stopping times.
along the path as the upper bound of the volatility integral. And for each maturity all strikes are calculated based on the same integral.

The use of Monte Carlo simulation will introduce numerical error, stemming from the sampling of the volatility distribution. The numerical error will converge towards zero as the number of volatility paths is increased. But as we include the simulation in the iterative search for the minimum of the penalty function, the computation time is of great concern. We therefore analyze carefully the numerical error of the calibration model, but as additional convergence issues arise when utilizing the model for pricing, the analysis has been postponed to Section 11.1. As we will argue in that section an acceptable level of error is obtained by 8 time steps per year and 2048 volatility paths using low discrepancy numbers as described above. This choice of time steps and volatility paths leads to a computation time of approximately 50 seconds for each evaluation of the penalty function. Remembering that each iteration in the minimization requires calculation of partial derivatives with respect to all the parameters, calibration of the model is a heavy computational burden.

In the extended model there are 17 parameters to be determined. For each volatility coefficient $a$, $b$, $\ln c$ and $\ln d$ one must determine the reversion speed, $RS$, the reversion level, $RL$, the volatility in the Ornstein Uhlenbeck process, $\sigma$ or $\sigma$, and the initial level of the coefficients, $\text{Initial}$, plus the displacement coefficient, $\alpha$ or $\alpha$. It would be possible to just minimize the penalty function over the 17 parameters, but some issues make this procedure unappealing. First the number of parameters is overwhelming, and the concern for the model being over-parameterized is justified. By extending the concept of time-homogeneity we could argue that time-homogeneity could be imposed by restricting the reversion levels to be equal to the initial value of the coefficients, i.e. imposing the restriction:

\begin{equation}
RL_i = \text{Initial}_i \quad \text{for } i = a, b, c \text{ and } d
\end{equation}
This would be financially justifiable as long as we assume that there is nothing special about the instantaneous volatility function as of today.\(^49\) Thereby we reduce the number of parameters available for fitting the model to 13.

This is still a large number of free parameters, but Joshi & Rebonato (2003) suggests that if one is uncomfortable with the large number of parameters one can keep \(a, b\) and \(\ln c\) constant and let the stochastic volatility be driven exclusively by a stochastic \(\ln d\). This corresponds to setting \(\sigma_a = \sigma_b = \sigma_c = 0\) and thereby making the corresponding RS irrelevant.

We choose to impose this restriction, thereby reducing the number of parameters to 7. We do this as our calibration results showed that the caplets we calibrate to contained very little information which could determine the parameters of these three processes. This was discovered as large changes in these parameters lead to very small changes in the penalty function. So the same caplet prices could be produced from very different values of \(\sigma_a, \sigma_b, \sigma_c, \text{RS}_a, \text{RS}_b\) and \(\text{RS}_c\). Though the caplet prices do not depend very much on the value of these parameters, it is very possible that the value of exotic products would. But as we have no information to determine the correct parameters of the processes we consider \(a, b\) and \(\ln c\) as deterministic. The stochastic volatility is thereby exclusively induced by \(\ln d\) following an Ornstein Uhlenbeck process.

So our calibration procedure consists of determining the 7 parameters of the restricted model.\(^50\)

To facilitate an analysis of the separate effects of the displaced diffusion and the stochastic volatility, we will examine special cases, which allow only one of these effects. Also we calibrate a deterministic volatility model in this calibration setup. This model corresponds exactly to the standard model, but calibrating and pricing in the same setup as the other models allows a direct comparison. This results in four models: a deterministic volatility model, DET, a displaced diffusion model, DD, a

\(^{49}\) Otherwise if we believe that the volatility function will revert to some other form, we could equally impose this on the reversion level coefficients

\(^{50}\) The calibration code and the following pricing code is build to facilitate an unrestricted model
stochastic volatility model, SV, and the full displaced diffusion, stochastic volatility model, DDSV. The models are obtained by imposing the following additional restrictions on the model:

**Table 5: Definition of extended models**

<table>
<thead>
<tr>
<th>Model</th>
<th>Restrictions imposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET</td>
<td>( \sigma_d = 0 )</td>
</tr>
<tr>
<td>DD</td>
<td>( \alpha = 0 )</td>
</tr>
<tr>
<td>SV</td>
<td>( \sigma_c = 0 )</td>
</tr>
<tr>
<td>DDSV</td>
<td>( \alpha = 0 )</td>
</tr>
</tbody>
</table>

The documentation contains a spreadsheet with a fully working calibration setup employing the build-in Excel Solver add-in. The calibration setup is built in Excel in order to show in an intuitive way the minimization. But the actual results reported are obtained using a VBA based calibration setup using Frontline Systems Solver DLL.\(^{51}\) The VBA environment and the improved solver allows a considerably lower calibration time.

Even using this more efficient calibration setup the minimization takes between 2-8 hours depending on the starting parameters. As compared to the standard model, this is an enormous increase, and it means that it is not possible to recalibrate the model during a trading day using our implementation. The model is therefore not suitable as a trading model, but could possibly be utilized in the context of analysis. One possibility would be as a benchmark for other models e.g. the standard model, where one assumes that the extended model is a more correct model and the pricing errors of other models are calculated from the extended model. Another possibility would be in the context of risk analysis where a lower recalibration frequency could be acceptable. As our implementation has not been focused on efficiency it is very possible that the calibration time could be reduced considerably, but it is still a contrast to the simple and fast calibration of the standard model.

\(^{51}\) I would like to thank Jyske Bank for the use of the solver.
The results of the calibration procedure are given in the following table:

<table>
<thead>
<tr>
<th>Table 6: Volatility parameters (Extended models)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Deterministic</th>
<th>RS</th>
<th>RL</th>
<th>Sigma</th>
<th>Initial</th>
<th>Alfa</th>
<th>Penalty function</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>3,085</td>
<td>0</td>
<td>3,085</td>
<td>0</td>
<td>1,307</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>12,216</td>
<td>0</td>
<td>12,216</td>
<td>0</td>
<td>12,621</td>
</tr>
<tr>
<td>ln c</td>
<td>1</td>
<td>-0,675</td>
<td>0</td>
<td>-0,675</td>
<td>0</td>
<td>-0,453</td>
</tr>
<tr>
<td>ln d</td>
<td>1</td>
<td>2,292</td>
<td>0</td>
<td>2,292</td>
<td>0</td>
<td>1,912</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stochastic Volatility</th>
<th>RS</th>
<th>RL</th>
<th>Sigma</th>
<th>Initial</th>
<th>Alfa</th>
<th>Penalty function</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>-1,484</td>
<td>0</td>
<td>-1,484</td>
<td>0</td>
<td>1,058</td>
</tr>
<tr>
<td>ln c</td>
<td>1</td>
<td>23,239</td>
<td>0</td>
<td>23,239</td>
<td>0</td>
<td>14,780</td>
</tr>
<tr>
<td>ln d</td>
<td>1</td>
<td>-0,381</td>
<td>0</td>
<td>-0,381</td>
<td>0</td>
<td>-0,330</td>
</tr>
<tr>
<td>ln d</td>
<td>1,174</td>
<td>1,957</td>
<td>0,888</td>
<td>1,957</td>
<td>0,817</td>
<td>0,984</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Displaced Diffusion</th>
<th>RS</th>
<th>RL</th>
<th>Sigma</th>
<th>Initial</th>
<th>Alfa</th>
<th>Penalty function</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>0,193</td>
<td>0</td>
<td>0,193</td>
<td>0,018</td>
<td>0,872</td>
</tr>
<tr>
<td>ln c</td>
<td>1</td>
<td>12,621</td>
<td>0</td>
<td>12,621</td>
<td>0,032</td>
<td>0,435</td>
</tr>
<tr>
<td>ln d</td>
<td>1</td>
<td>-0,463</td>
<td>0</td>
<td>-0,463</td>
<td>0</td>
<td>-0,463</td>
</tr>
<tr>
<td>ln d</td>
<td>1</td>
<td>1,912</td>
<td>0</td>
<td>1,912</td>
<td>0</td>
<td>1,912</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Displaced Diffusion Stochastic Volatility</th>
<th>RS</th>
<th>RL</th>
<th>Sigma</th>
<th>Initial</th>
<th>Alfa</th>
<th>Penalty function</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>-0,825</td>
<td>0</td>
<td>-0,825</td>
<td>0</td>
<td>0,435</td>
</tr>
<tr>
<td>ln c</td>
<td>1</td>
<td>14,780</td>
<td>0</td>
<td>14,780</td>
<td>0</td>
<td>14,780</td>
</tr>
<tr>
<td>ln d</td>
<td>1</td>
<td>-0,330</td>
<td>0</td>
<td>-0,330</td>
<td>0</td>
<td>-0,330</td>
</tr>
<tr>
<td>ln d</td>
<td>0,817</td>
<td>0,984</td>
<td>1,190</td>
<td>0,984</td>
<td>0,817</td>
<td>0,984</td>
</tr>
</tbody>
</table>

The first observation we make is that, although the deterministic calibration corresponds to the standard model calibrated above in Section 10.3.1, the resulting coefficients are not equal. This is because the standard model is only calibrated to 4% caps, while the DET model is calibrated to the full volatility surface. Of course if we wanted we could reduce the calibration sample to a single strike which would make the calibration directly comparable with the standard model calibration. We still do not expect an exact reproduction of the parameters because of differences in the calibration procedure, mainly in the calculation of the volatility integrals. As we will use the DET implementation of the standard model in the following comparison, we base the calibration on the full calibration sample in order to make the calibration comparable to the other models.

Of course we observe that the fit to the caplet implied volatility surface is improved as we include more and more parameters in the calibration. We also see that introducing the displaced diffusion effect allows a lower mean square error than including stochastic volatility effect. This indicates that the skew inducing feature is superior in terms of fit to the smile inducing.

In addition we see that the improvement in fit for the DDSV model when going from the DET model is greater than the sum of the improvements for DD and SV. This is because the DDSV calibration results in a larger skew which allows a closer fit to the ITM caplets. This is made possible by the smile inducing part of the DDSV. The smile compensate for the larger skew, which reduces the fit of the OTM caplets, by increasing the implied volatilities for these caplets.
When looking at the resulting implied volatility surfaces we clearly see the effects induced by the different models. The DET model produces a volatility surface that is flat across strikes. We see a small divergence from this as the highest strike for the smallest maturity apparently experiences a higher implied volatility than the other caplets of the same maturity. This is merely a computational issue. The implied volatility is obtained from the caplet premium by a search algorithm that changes the implied volatility until the resulting caplet premium corresponds to the desired value with some accuracy. For the particular caplet the vega is very low, meaning that the sensitivity of the option price with respect to changes in the implied volatility is very low. It is therefore possible to get within the desired accuracy of the caplet premium, even if the implied volatility is quite some distance away from the “true” value. We observe this result for all four models.52

52 The problem could be remedied by reducing the accuracy in the implied volatility search or excluding the problematic caplet from the penalty function.
When moving to the DD model we see that the implied volatility surface exhibits a nice skew. The best fit implied volatility for the SV model exhibits a pronounced smile for the shortest maturities, and a very shallow smile for longer maturities. The model volatility thereby fit the general shape of the market volatility surface, which exhibits a large smile for the short maturities and only a shallow smile for the longer maturities. Finally the combined model of course displays a combination of skew and smile.

**Figure 14:** Calibration errors  
(Extended models)

Figure 4 shows the differences between the market and model implied volatility surfaces. The systematism in the errors from the interpolation scheme observed in the standard model calibration is also observed here. And the characteristics of each of the models just presented are also observable in these representations. We see that the DET and the SV models leave a considerable skew in the errors, which is evidence of a misspecification of the models. This skew is greatly reduced in the DD and DDSV models. But we see that some skew is still left in the errors for the long maturities,
which is evidence of an underpricing (overpricing) of OTM (ITM) caplets. What is not directly observable from the chosen graphs\textsuperscript{53} is that the errors of the short maturities exhibit a tendency of an opposite skew; ITM (OTM) caplets are underpriced (overpriced). This feature of the error surface means that the model does not adequately account for the high implied volatilities of short OTM options.

For the long maturities we see that a large part of the remaining error is due to the interpolation, and it is expected that the fit could be improved considerably if the market volatility surface was smooth. The observation also highlights the call for a parameterized approach to smooth the market volatility.

Another observation we make is that the short maturities exhibit larger errors than the longer maturities. This was expected as the market volatility is less smooth in the shorter than the longer end. We recall that the difficulty in fitting especially the short end was also experienced in the yield curve calibration. It could therefore be a subject of further analyzes to investigate if the pricing errors in the short end are a result of the calibration procedure for the yield curve or from the inability of the model to capture some subtleties of the short end.

What is not evident from these results is that the calibration result depended greatly on the starting parameters. It was therefore necessary to repeat the calibration for different sets of starting parameters in order to investigate if the previously obtained results represented the global minimum of the penalty function or just a local one. In a practical implementation with regular recalibrations one would probably use as starting parameters the ones obtained from the latest calibration, which would prove successful if the market volatility only changed little from calibration to calibration.

But if the basic shape of the volatility surface would change between recalibrations one would risk the possibility of obtaining a local minimum instead of the global. As an alternative or supplement to this and in the situations where no previous parameters are available one must qualitatively asses the market volatility surface and choose starting parameters based on this. A first guess for the initial values and thereby the reversion levels of the coefficients could be obtained from a calibration of the standard model, remembering that the displaced diffusion and stochastic volatility would reduce the reversion level of the instantaneous volatility necessary to produce

\textsuperscript{53} The observation can be made from the spreadsheets in Volatility.xls
the same implied volatility. The displacement coefficient could then be set in correspondence to the magnitude of the skew. Finally the stochastic volatility parameters, which proved to be the hardest to obtain global best fit values for, could be determined from the characteristics of the smile. The more pronounced the smile the higher the sigma in the process for ln d, and the difference in the magnitude of the smile across maturities is controlled by the reversion speed. The greater the reversion speed the more shallow the smile for long maturities. These parameters, resulting in the desired qualitative shape of the model implied volatility surface, could then be used as starting point for the calibration.

But the need for repeated calibrations each with a computation time of several hours reduces the practical relevance of the model. A practical implementation of the model requiring daily recalibrations would therefore require an improved minimization technique, which improves the probability of obtaining the global minimum. Such procedures exist but often come at the cost of an increased computation time. The improved minimization technique would therefore require a reduction of the computation time for each iteration. I believe the greatest improvements of the computation time can be obtained by improving the sampling of the TOTC matrix.

The final judgement of the plausibility of the obtained parameters must be based on an analysis of the stability of the obtained parameters when repeating the calibration on different days. If the dynamics are modelled correctly by the DDSV, we would expect the parameters to be fairly constant over time. This analysis is outside the scope of this paper, but should be a subject of future research in order to validate the model specification.

The analysis shows that the extended model affords a considerably better fit than the standard model represented by the DET model. We also observe that the combination of a skew and a smile affords a significant increase in fit over models allowing only one of the effects. This is because the smile allows the possibility of a larger skew without being penalized by a worsening of the fit for short maturity OTM caplets. This better fit unfortunately comes at the cost of an enormous increase in the calibration time, and a considerable possibility of obtaining a local minimum of the penalty function instead of the global. This means that the extended model in the current implementation is not suitable for trading purposes requiring a frequent
recalibration of the model. As the efficiency of the calibration procedure has not been a major concern in this paper it is very possible that the calibration time could be reduced considerably, mainly by a more efficient sampling of the volatility and thereby faster convergence, allowing the calibration to be based on fewer volatility paths.
11 Convergence

For all problems solved by numerical methods the issue of convergence must be addressed. Numerical methods provide us with an estimate of the value of the product evaluated, but this estimate is subject to numerical error. The numerical error can be reduced by increasing the accuracy of the specific method at the cost of a longer computation time.

For a Monte Carlo simulation we can estimate the integral more accurately by sampling more paths, which will of course increase the computation time. So the analysis of convergence is about balancing accuracy and computation time, as to obtain an estimate that we are confident is not too far from the true value without incurring too great a computational burden.

For our problem three sources of numerical error exist: The number of volatility paths used, i.e. the number of different TOTC matrices sampled, the number of forward rate paths used to estimate the value of the products given each TOTC matrix and the discretization error of the volatility processes. The discretization error is typically not considered as a convergence problem, but we will analyse it in this section as the procedure resembles the procedure used for the traditional convergence issues. The discretization error depends on the length of the time steps employed in the evolution of the volatility coefficients, i.e. the number of time steps per year.

As the convergence will depend on the product being evaluated, we will conduct the analysis in two parts. One part will analyse the convergence of standard cap prices for the purpose of determining the number of volatility paths and time steps when calibrating the model. As the calibration uses an exact analytical solution for the cap prices given the volatility path, it is not necessary to investigate the forward rate paths dimension in this part of the analysis. Another part will analyse convergence of the pricing model including the number of time steps per year, volatility paths and forward rate paths. The analysis is based on the DDSV model with parameters obtained from a preliminary calibration of the model presented in Appendix 6: Preliminary parameters used in the convergence analysis. The parameters are not exactly identical to the final parameters of the DDSV model. The convergence could

---

54 The analysis will not consider the accuracy of the Predictor Corrector approximation
depend on the particular parameters, but we assume that the results carry over unaltered to the new parameters.

Before we proceed with the actual analysis we will consider some general convergence subjects and introduce the procedure of the analysis.

For the choice of the number of time steps per year in the volatility evolution, the discretization error will go towards zero as we increase the number of time steps per year. In order to decide on an acceptable number of time steps we calculate an “exact” option value using a high number of time steps. We then calculate the option value using a lower sampling frequency and compare this to the “exact” value. The calculations are set up so the volatility evolution is kept fixed. For each price the volatility is evolved over the same number of time steps as in the “exact” calculation using the same set of random numbers, making the volatility evolution exactly equal. Based on this volatility evolution the TOTC matrix is constructed by sampling the volatility by different frequencies, i.e. using different time steps per year. This way we can isolate the discretization error from numerical error.

Concerning the issue of convergence, the Monte Carlo estimate of the price will depend on the particular set of random numbers drawn, and a new set of random numbers will result in a different price. So the Monte Carlo estimate will fluctuate around the true value of the product. The difference between the true value and the Monte Carlo estimate is the numerical error. According to the Law of large numbers the standard Monte Carlo estimator will converge towards the true value of the product\textsuperscript{55} as we increase the number of paths, i.e. in the limit as the number of Monte Carlo paths goes to infinity, the Monte Carlo estimator of the value goes towards the true value of the product being evaluated. But as we are only evaluating a finite (and preferable small) number of paths we must analyse the numerical error incurred. If we have an exact analytical solution of the price we can compare the Monte Carlo estimate to the exact price and thereby calculate the error, but as an analytical solution is not always available, and the distribution of the errors is also of interest we will analyze the standard error of the Monte Carlo estimate.

\textsuperscript{55} Given the assumptions underlying the model
In a standard Monte Carlo simulation the standard error of the Monte Carlo estimator is given by\(^{56}\)

\[
se_{MC} = \sqrt{\frac{\sigma^2(f)}{N}}
\]

(65) 

where 

\(\sigma^2(f)\) is the variance of the function evaluated by Monte Carlo 

\(N\) is the number of paths

The variance of the function can be estimated by the standard deviation of the realisations of the paths. We see that the standard error of the standard Monte Carlo estimator scales by \(1/\sqrt{N}\), the rate of convergence is of \(O(\sqrt{N})\). So in order to reduce the standard error by a factor 10 it is necessary to increase the number of iterations by a factor 100.

As we are combining Quasi Monte Carlo, for the volatility paths, and standard Monte Carlo, for the forward rate paths, the standard error of the estimator is no longer given by Equation (65). But it can be shown that in optimal cases the rate of convergence can be reduced to \(O(N^{-1})\), when using low discrepancy numbers, which is much faster than the rate of convergence for the standard Monte Carlo.\(^{57}\)

For the same reason we cannot use Equation (65) as a measure of convergence. We therefore take a more qualitative approach. Our goal is still to assess the dispersion of the estimate and to decide on a number of paths that ensures confidence in the estimate.

We proceed by calculating 10 estimates of the value using each time a different set of random numbers. The dispersion of the estimates can then be evaluated by calculating different statistics of the estimates and plotting the 10 prices.

Using only 10 realizations of the prices means that the dispersion statistics are not very accurate estimates of the population parameters. But as we will see later the accuracy suffices for the purpose of our analysis, and great intuition can be reaped from the procedure.

---

\(^{56}\) A result of the Central Limit theory

\(^{57}\) Dias
11.1 Convergence of calibration setup

For the calibration the purpose of the convergence analysis is to ensure that the parameters obtained by the calibration are not specific to the particular set of random numbers underlying the calibration. We need to be certain that another set of random numbers would not produce prices that are very different from the prices from the current random numbers, as this would lead to a different value of the penalty function, and we would not be certain that this value would be the minimum. By using enough volatility paths to ensure that the prices and thereby the penalty function are invariate to a new draw of random numbers, we ensure that the calibration is not specific to the set of random numbers.

But first we analyze the discretization error as presented above.

The specific product analyzed is a 5 year standard cap strike 4% and the “exact” price is calculated using 40 time steps per year, which corresponds to a time step of 9 days.

<table>
<thead>
<tr>
<th>Number of time steps per year</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>20</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cap premium</td>
<td>162.32</td>
<td>161.82</td>
<td>161.65</td>
<td>161.62</td>
<td>161.57</td>
<td>161.54</td>
</tr>
<tr>
<td>Difference</td>
<td>0.78</td>
<td>0.28</td>
<td>0.11</td>
<td>0.08</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>%-difference</td>
<td>0.48%</td>
<td>0.17%</td>
<td>0.07%</td>
<td>0.05%</td>
<td>0.01%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Note: Standard cap, Time to maturity 5, Strike 4

Difference and %-difference is calculated relative to the 40 time steps per year price

We see that the discretization error measured as the deviation from the “exact” value is of course declining as we increase the number of time steps. But also we see that the rate at which it declines is falling. So the marginal benefit of increasing the number of time steps is decreasing.

As we find an discretization error below 0.10% acceptable, we conclude that 8 time steps per year corresponding to a time step of one and a half month is satisfactory for our analysis.

We then proceed to the analysis of the number of volatility paths.

58 Alternatively we could have analyzed directly the convergence of the penalty function.
The graph clearly illustrates the convergence issue. Using 64 volatility paths we would get prices ranging from 154 to 169, meaning that we would not be confident in the prices we obtain. Also we see clearly that as we increase the number of volatility paths, the dispersion of the estimate is reduced.

Accepting a standard deviation of 0.10 we choose to work with 2048 volatility paths. This is a relatively large uncertainty, but as the calibration is iterating over the price it is very important, that the computation time is kept at a minimum. The 2048 volatility paths and 8 time steps results in a computation time of each iteration around 50
seconds, which means that the calibration time is 2-8 hours depending on the starting parameters. The 2048 volatility paths is a stark contrast to the 64 volatility paths that Joshi & Rebonato (2003) report is sufficient to obtain an acceptable degree of convergence. Still, using 64 volatility paths in our implementation, each iteration would take around 3 seconds resulting in a calibration time up to half an hour, which is still much longer time than the standard model. A calibration time of half an hour would allow recalibrations during a trading day, but it would still be insufficient for trading purposes.

The practical usefulness of the model would therefore depend greatly on the possibility of reducing the necessary number of volatility paths by a using a more efficient sampling of the volatility.

11.2 Convergence of pricing setup

Moving to the pricing model the goal of the convergence analysis is to determine the number of time steps per year, volatility paths and forward rate paths per volatility so that we are confident that the estimates obtained are close to the true value of the product.

The convergence analysis has been performed for both the Sticky cap and the Ratchet cap. Because the results are very similar we only present the results for the Sticky cap. The particular product investigated is a 5 year Sticky cap, margin 0,25%.

Our analysis of the discretization error confirms the conclusion above that 8 time steps per year is sufficient to avoid bias in the results due to the discretization of the volatility process, therefore we do not reproduce the results.

In order to analyze the separate effect of volatility paths and forward rate paths we first keep the random numbers driving the forward rate fixed and investigate the number of volatility paths, second we keep the volatility evolution invariate by fixing the number of volatility paths and the random numbers driving the volatility and investigate the number of forward rate paths per volatility path. By analysing forward rate paths per volatility paths instead of total number of forward rate paths we achieve that adding volatility paths does not alter the convergence of the prices from each volatility path. So the final prices, which can be seen as averages of the individual
volatility paths, are comparable as the prices from each volatility path are calculated in the same way, i.e. based on the same number of forward rate paths.

The following figure and table shows the convergence for the number of volatility paths.

**Figure 16:** Convergence for volatility paths  
(Pricing setup Sticky cap)

![Convergence for volatility paths](image)

Note: The graph plots for each number of volatility paths 10 prices for a 5Y Sticky cap, margin 0.25% using for each price a new set of random numbers for the volatility evolution while keeping the random numbers in the forward rate evolution fixed.

**Table 9:** Convergence for volatility paths  
(Pricing setup Sticky cap)

<table>
<thead>
<tr>
<th>Number of volatility paths</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
<th>2048</th>
<th>4096</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>351.0</td>
<td>238.1</td>
<td>86.6</td>
<td>67.6</td>
<td>27.8</td>
<td>14.2</td>
<td>11.0</td>
</tr>
<tr>
<td>Range</td>
<td>1.015</td>
<td>697.0</td>
<td>236.4</td>
<td>201.9</td>
<td>87.4</td>
<td>43.6</td>
<td>32.8</td>
</tr>
<tr>
<td>Range normalized</td>
<td>11.18%</td>
<td>7.80%</td>
<td>2.59%</td>
<td>2.22%</td>
<td>0.96%</td>
<td>0.48%</td>
<td>0.36%</td>
</tr>
</tbody>
</table>

Note: Range normalized is the range normalized by the average of the estimated prices

The statistics are for 10 prices of a 5Y Sticky cap, margin 0.25% using for each price a new set of random numbers for the volatility evolution while keeping the random numbers in the forward rate evolution fixed.
The prices are calculated using 8 time steps per year and 2000 forward rate paths per volatility path. Again we see the typical convergence pattern. We see that we must use at least 1024 volatility paths in order to obtain an acceptable degree of convergence.

**Figure 17:** Convergence for forward rate paths  
(Pricing setup Sticky cap)

Note: The graph plots for each number of forward rate paths 10 prices using for each price a new set of random numbers for the forward rate evolution while keeping the random numbers in the volatility evolution fixed.

<table>
<thead>
<tr>
<th>Number of forward rate paths per volatility path</th>
<th>250</th>
<th>500</th>
<th>750</th>
<th>1000</th>
<th>1500</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>10,7</td>
<td>5,3</td>
<td>3,1</td>
<td>1,9</td>
<td>1,1</td>
<td>1,1</td>
</tr>
<tr>
<td>Range</td>
<td>32,5</td>
<td>16,4</td>
<td>11,1</td>
<td>6,4</td>
<td>3,6</td>
<td>3,5</td>
</tr>
<tr>
<td>Range normalized</td>
<td>0,36%</td>
<td>0,18%</td>
<td>0,12%</td>
<td>0,07%</td>
<td>0,04%</td>
<td>0,04%</td>
</tr>
</tbody>
</table>

Note: Range normalized is the range normalized by the average of the estimated prices  
The statistics are for 10 prices of a 5Y Sticky cap margin 0,25% using for each price a new set of random numbers for the forward rate evolution while keeping the random numbers in the volatility evolution fixed.

The prices are calculated using 8 time steps per year and 2048 volatility paths.  
We see that the convergence pattern for the forward rate paths per volatility path is similar to the pattern for the volatility paths. But we see that the magnitude of the dispersion is much smaller.
Based on this analysis we choose 2048 volatility paths and 500 forward rate paths for each volatility path. This results in a standard deviation of 19.1, a range of 63.8 and a range normalized of 0.70% for 10 prices, when for each price a new sample of random numbers are drawn for both the volatility and the forward rate evolution. This results in a calculation time around 12 minutes for a 10 year option. We have thereby obtained an acceptable but not overwhelming accuracy, but we do not increase the number of paths additionally due to the already long computation time. Instead we ensure that the prices in the comparison are calculated using exactly the same forward rate evolutions, by employing the same set of random numbers. Thereby the numerical error will affect all prices similarly and conclusions regarding pricing differences between models can be considered reliable.
12 Pricing

Having implemented the models we now turn to a comparison between the prices the models produce.

To ensure that the conclusions are not distorted by differences in implementation and numerical error, we utilize the DET calibration of the standard model, while we remember that this calibration constitute an average across strikes. We also price all products in the exact same model differing only in the payoff function, and using the same sets of random numbers for all calculations.

The main interest in our analysis is the Ratchet and the Sticky caps, as these products depend on the terminal correlation. But we will also include in the analysis a Standard cap. The cap rate is chosen as 4%, making the cap approximately ATM. The Standard cap does not depend on the terminal correlation, so any pricing difference for the Standard cap will stem from the calibration. This reminds us that the pricing differences between the DET and the other models will reflect partly differences due to changes in terminal correlation and partly differences due to the fact that the average volatility in the DET model is not representative of the underlying process, and will result in biased prices for some product characteristics.

The goal of this section is not to determine the best model. This would require market prices of the two exotic caps in order to compare the model prices to the true value proxied by the market value. Also it would require a comparison across time in order to evaluate the general performance, not restricted to one particular calibration day. Rather the section will concentrate on highlighting the differences in the prices from the different models as a continuation of the general comparison of the models.

Taking a first look at the pricing by examining the prices for 10 year caps with a margin of 25 bp for the Ratchet and the Sticky cap, we see the following.
Table 11: Comparison of products

Prices and percentage changes from the DET model

<table>
<thead>
<tr>
<th></th>
<th>Sticky</th>
<th>Ratchet</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET</td>
<td>23.506</td>
<td>8.380</td>
<td>45.970</td>
</tr>
<tr>
<td>DD</td>
<td>23.886</td>
<td>8.109</td>
<td>46.789</td>
</tr>
<tr>
<td>SV</td>
<td>23.845</td>
<td>7.883</td>
<td>45.385</td>
</tr>
<tr>
<td>DDSV</td>
<td>24.077</td>
<td>7.813</td>
<td>45.286</td>
</tr>
</tbody>
</table>

Note: Time to maturity 10 years, Exotic margin 25 bp, Standard caprate 4.0%, NP 1.000.000
The percentage change is the pricing difference between each model and the DET model

The percentage change from DET is the pricing difference between each model and the DET model.

We see that the prices across models differ up to 6.77%. This confirms that the discussion of which model to use is relevant.

We also see that the pricing differences must be analyzed individually for each product we price. The Sticky and the Ratchet cap which has some features in common does not exhibit the same pricing differences across models, actually the more complex models increase the price of the Sticky, while they decrease the price of the Ratchet.

For a standard cap the DET model would underprice the OTM and ITM as compared to the ATM caplets as the model does not account for the non flat volatility leading to a mispricing of caplets with certain characteristics/strikes. We therefore also expect the pricing differences to depend on the characteristics of the exotic caps. We will consider changes in the margin and the time to maturity.
First we see that the price decreases for a higher margin as expected from the properties of the products.

Additionally we now see that the percentage change when going from DET to DD is falling for a higher margin. We see that a high margin for the Sticky cap will result in the DD price being lower than the DET price as is observed for all margin levels for the Ratchet. This is a natural relationship. Increasing the margin is similar to increasing the cap rate of a Standard cap making it OTM. For the DD the skewness results in lower prices for OTM caps as compared to a deterministic volatility model. The Sticky and Ratchet caps exhibit the same property that increasing the margin, which ceteris paribus will increase the strike of each caplet, reduces the price in a skewed model more than in the flat volatility model.

The same line of reasoning can be applied to the SV model for the Sticky cap, where percentage change is increasing for a higher margin. In the presence of stochastic volatility, the kurtosis makes the OTM standard cap more valuable as compared to the DET value, which translates into the increasing %-change for a higher margin observed for the Sticky cap. For the Ratchet this effect is not observed. This is because the constant resetting of the cap rate without the minimum function of the Sticky cap makes each realization of the caplets more ATM than compared to the Sticky, hence reducing the kurtosis effect. But we see some effect on the price which must instead stem in part from the changed terminal correlation.

As the DDSV allows a combination of the two effects we see that for an increasing margin level the pricing differences are smaller than for the two other models because of canceling of the two effects.
Table 13: Time to maturity analysis

Prices and percentage changes from the DET model

<table>
<thead>
<tr>
<th>Sticky Ratchet Standard</th>
<th>Time to maturity</th>
<th>Prices</th>
<th>Percentage Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>5</td>
<td>10</td>
<td>Model</td>
</tr>
<tr>
<td>DET</td>
<td>8.899</td>
<td>23.506</td>
<td>8.380</td>
</tr>
<tr>
<td>SV</td>
<td>9.373</td>
<td>23.848</td>
<td>8.733</td>
</tr>
<tr>
<td>DDSV</td>
<td>8.566</td>
<td>24.077</td>
<td>-1.86%</td>
</tr>
</tbody>
</table>

Note: Margin 25 bp, NP 1.000.000

The percentage change is the pricing difference between each model and the DET model.

We see that the pricing differences also depend on the time to maturity. We see that for the 5Y Sticky cap the effect of the SV and DDSV is of opposite sign, and for the Ratchet the pricing difference between DET and DDSV is increased to 9.82%. This again highlighting that in order to asses the pricing differences of a product we cannot apply inferences from analysis of other products or even the same products with other characteristics. We must analyze individually the pricing differences of each product of interest.

We have now analyzed some pricing differences, when we change some characteristics of the product. We will now examine the sensitivities of the prices to changes in the underlying parameters. These sensitivities would determine the hedging strategy for the issuer of the products and are therefore often of greater concern for the issuer than the price itself.

Assuming the issuer was interested in hedging completely the price of the products she should build a portfolio, which did not change value for any change in variables or parameters. Recalling the calibration procedure, this would mean a portfolio value invariate to changes in the yield curve, volatility, correlation and time.

For interest rate products yield curve changes will affect the value by changing both the value of the underlying and the numeraire. Volatility changes would result in a changed implied volatility surface translating into changed volatility parameters and displacement coefficient. As the correlation is not observable, a change in correlation would be introduced as a revision of the correlation parameter in Equation (56). As time goes by the time to maturity of the products underlying the portfolio would
change and hence the portfolio value would change. Discarding the analysis of time, we should analyze the changes in the yield curve, volatility and correlation.

In the deterministic model the volatility could be hedged by making the value invariate to changes in the parameters of the deterministic volatility function. As these parameters holds a financial interpretation as determining the shape of the instantaneous volatility function and is closely linked to the Black implied volatility, this hedging approach is highly intuitive. For instance will a portfolio invariate to changes in d be invariate to approximately parallel shifts in the implied volatility. This approach would hedge changes in volatility due to both previously falsely determined parameters and actual changes in the volatility. As a change in the parameters of the deterministic volatility would result in a direct change in the TOTC matrix, the procedure is quite intuitive.

This is not the case for the displaced diffusion and stochastic volatility models. For the stochastic volatility the problem is that the coefficients of the volatility function is not directly determined by the calibration. We determine the current value for the coefficients and the parameters of their processes, which in turn determines the possible realization of the TOTC matrix, the distribution of which determines the value of our products. Hedging volatility in the stochastic volatility would then mean that we should hedge changes in the TOTC distribution stemming from both the inherent stochasticity of the volatility and the possible changes in parameters. As the volatility is no longer determined directly by the coefficients of the volatility function, we must be very careful in the way we define our hedging parameters. For instance we restricted the reversion levels of the coefficients to be equal to the current level of the coefficients. If we maintain this restriction, hedging with respect to a volatility coefficient would be a hedge against a simultaneous move in the coefficient and the reversion level. Alternatively we could loosen this restriction letting the reversion level be fixed and hedge only the stochastic change in the coefficient.

For the displaced diffusion the simple connection between the parameters of the volatility function and the TOTC matrix is also lost. The approximately parallel shift in the implied volatility surface is not attainable as a simple parallel shift in the instantaneous volatility function.
As these aspects of volatility hedging now briefly discussed requires considerably analysis I will leave it for further research.

Instead I will focus on sensitivities to changes in the yield curve and the instantaneous correlation, which are more straightforward.

The analysis of yield curve changes is facilitated by the use of the Nelson-Siegel parametric function for the yield curve. Changes in the yield curve can be induced by shifting the Nelson-Siegel parameters which in turn will change both forward rates and the numeraire in a consistent manner. Also the parameterization allows an easy implementation of standard yield curve shifts. A parallel shift in the curve is obtained by shifting the $\beta_0$ parameter, while an almost linear slope shift is obtained by shifting $\beta_2$. In the interpretation of the sensitivities we must remember that the Nelson-Siegel yield curve estimated is for the zero coupon rate with annual compounding. The parallel and the slope shifts are therefore with respect to this curve and not the underlying forward rate. This is important to remember if the sensitivities are compared with to sources.

In the analysis the shifts have been applied after the first cap rate has been determined based on the unshifted spot rate.

The parallel shift is obtained by increasing $\beta_0$ by 1, this results in a 100bp parallel shift in the zero coupon rate with annual compounding resulting in an almost parallel shift in the underlying semi-annual forward rates of around 98bp. The slope shift is obtained by increasing $\beta_2$ by 8.575 and $\beta_0$ by -0.029. This results in an almost linear shift in the zero rates, where R(0,0.5) is unaltered and R(0,10) is increased by 50bp, which is equivalent to an increase in f(0,10) of approximately 200bp.

The shifted yield curves are illustrated in Appendix 7: Shifted yield curve and correlation surface.

---

59 The difference is due to differences in compounding.
### Table 14: Yield curve sensitivity analysis (Parallel shift)

<table>
<thead>
<tr>
<th></th>
<th>Sticky Ratchet</th>
<th>Current yield curve</th>
<th>Shifted yield curve</th>
<th>Difference</th>
<th>%-Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET</td>
<td>23.506</td>
<td>45.181</td>
<td>21.675</td>
<td>92.21%</td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>23.868</td>
<td>42.989</td>
<td>19.103</td>
<td>79.96%</td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>23.645</td>
<td>45.569</td>
<td>21.924</td>
<td>91.07%</td>
<td></td>
</tr>
<tr>
<td>DDSV</td>
<td>24.077</td>
<td>41.241</td>
<td>17.164</td>
<td>71.28%</td>
<td></td>
</tr>
</tbody>
</table>

|          | DET            | 8.380               | 14.275              | 5.895      | 70.34%       |
| DD       | 8.109          | 13.115              | 5.006               | 61.74%     |
| SV       | 7.883          | 13.114              | 5.227               | 68.73%     |
| DDSV     | 7.133          | 12.516              | 4.383               | 60.19%     |

Note: Time to maturity 10 years, Margin 25 bp, NP 1,000,000

The Difference and %-Difference is the absolute and relative difference when going from the current to the shifted yield curve.

The percentage change is the pricing difference between each model and the DET model.

|          | DET            | 45.970               | 94.396              | 48.426      | 105.34%      |
| DD       | 46.789         | 94.049               | 47.260              | 101.01%     |
| SV       | 45.385         | 94.223               | 48.838              | 107.61%     |
| DDSV     | 45.286         | 93.150               | 47.884              | 105.87%     |

|          | DET            | 8.380               | 14.275              | 5.895      | 70.34%       |
| DD       | 8.109          | 13.115              | 5.006               | 61.74%     |
| SV       | 7.883          | 13.114              | 5.227               | 68.73%     |
| DDSV     | 7.133          | 12.516              | 4.383               | 60.19%     |

Note: Time to maturity 10 years, Margin 25 bp, NP 1,000,000

### Table 15: Yield curve sensitivity analysis (Slope shift)

<table>
<thead>
<tr>
<th></th>
<th>Sticky Ratchet</th>
<th>Current yield curve</th>
<th>Shifted yield curve</th>
<th>Difference</th>
<th>%-Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET</td>
<td>23.506</td>
<td>35.083</td>
<td>11.577</td>
<td>49.25%</td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>23.868</td>
<td>34.174</td>
<td>10.289</td>
<td>43.07%</td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>23.645</td>
<td>35.239</td>
<td>11.584</td>
<td>47.76%</td>
<td></td>
</tr>
<tr>
<td>DDSV</td>
<td>24.077</td>
<td>33.058</td>
<td>8.981</td>
<td>37.30%</td>
<td></td>
</tr>
</tbody>
</table>

|          | DET            | 8.380               | 11.202              | 2.822      | 33.97%       |
| DD       | 8.109          | 10.427              | 2.318               | 28.58%     |
| SV       | 7.883          | 10.598              | 2.715               | 34.44%     |
| DDSV     | 7.133          | 9.819               | 2.680               | 25.68%     |

Note: Time to maturity 10 years, Margin 25 bp, NP 1,000,000

The Difference and %-Difference is the absolute and relative difference when going from the current to the shifted yield curve.

The percentage change is the pricing difference between each model and the DET model.

<table>
<thead>
<tr>
<th></th>
<th>Sticky Ratchet</th>
<th>Current yield curve</th>
<th>Shifted yield curve</th>
<th>Difference</th>
<th>%-Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DET</td>
<td>45.970</td>
<td>69.700</td>
<td>23.730</td>
<td>51.62%</td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>46.789</td>
<td>69.851</td>
<td>23.062</td>
<td>49.18%</td>
<td></td>
</tr>
<tr>
<td>SV</td>
<td>45.385</td>
<td>69.220</td>
<td>23.838</td>
<td>52.32%</td>
<td></td>
</tr>
<tr>
<td>DDSV</td>
<td>45.286</td>
<td>68.319</td>
<td>23.033</td>
<td>50.66%</td>
<td></td>
</tr>
</tbody>
</table>

|          | DET            | 8.380               | 11.202              | 2.822      | 33.97%       |
| DD       | 8.109          | 10.427              | 2.318               | 28.58%     |
| SV       | 7.883          | 10.598              | 2.715               | 34.44%     |
| DDSV     | 7.133          | 9.819               | 2.680               | 25.68%     |

Note: Time to maturity 10 years, Margin 25 bp, NP 1,000,000

The Difference and %-Difference is the absolute and relative difference when going from the current to the shifted yield curve.

The percentage change is the pricing difference between each model and the DET model.

For all three products we see that the higher interest rate means an increase in the cap values. This is a matter of course for the standard cap, but we see the same property with smaller magnitude for the Sticky and Ratchet cap, even though the cap rate adjusts to the interest rate level, and makes the increase in value smaller.

The Standard cap shows a smaller sensitivity to the yield curve shifts in the skewed model. This is because some of the increase in value due to a higher value of the underlying is offset by a decrease in the volatility, because of the less than proportional scaling of volatility with the underlying.
For the standard cap the stochastic volatility has a very small effect on the sensitivity. One would expect that this effect will be considerably higher if one applied a larger yield curve shift, making the option deep ITM. Going to the exotic options we see that the sensitivities are much more affected by a model change indicating that the joint distribution of rates is altered to a great degree when going from one model to the other.

In the current framework it is just as easy to analyze the sensitivity to changes in the instantaneous correlation. The instantaneous volatility is assumed to depend on only one parameter. We can thereby analyze the effects of changing the instantaneous correlation by shifting this parameter. When examining Equation (56) we see that increasing the parameter makes the correlation steeper as a function of the time between forward rates, i.e. reduces the instantaneous correlation. The shift in the instantaneous correlation is obtained by shifting beta by 0.005. This results in a steeper correlation function and the lowest correlation between maturity 0.5 and 10 years is reduced from 0.83 to 0.79. The shifted correlation surface is illustrated in Appendix 7: Shifted yield curve and correlation surface.

The standard cap value does not depend on the instantaneous correlation, and is therefore not included in this analysis.

### Table 16: Instantaneous correlation sensitivity analysis

<table>
<thead>
<tr>
<th>Prices and percentage changes from the DET model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sticky Correlation Analysis</td>
</tr>
<tr>
<td>DET</td>
</tr>
<tr>
<td>Current correlation</td>
</tr>
<tr>
<td>Shifted correlation</td>
</tr>
<tr>
<td>Difference</td>
</tr>
<tr>
<td>% Difference</td>
</tr>
<tr>
<td>DET</td>
</tr>
<tr>
<td>DET</td>
</tr>
<tr>
<td>DD</td>
</tr>
<tr>
<td>BV</td>
</tr>
<tr>
<td>DDSV</td>
</tr>
</tbody>
</table>

The table illustrates that the value of both products increase, when the instantaneous volatility is reduced. This is natural as the probability of encountering first a very low

---

60 We keep in mind that the instantaneous correlation is not the only source of correlation between forward rates (terminal correlation) cf. Section 8.2.1
forward rate, setting the strike at a corresponding low level, followed by a high
forward rate is higher for a lower correlation.
We see that the sensitivity is higher for the DD model and lower for the SV model as
compared to the DET model.
Also we see that although the correlation shift can be regarded as quite large the
resulting changes in values are rather small. Although we have not analyzed
sensitivities for other correlation shapes the results assure us that a precise estimation
of the instantaneous correlation is not vital for the pricing of the Sticky and Ratchet
caps analyzed.

This section has showed that the four models imply different prices for the products
analyzed. We have seen that the differences in prices depend on the characteristics of
the product. This was expected from the shape of the implied volatility surface as it
reveals that the deterministic model implying a flat volatility across strikes does not
result in the right prices for some product characteristics, i.e. ITM options. Though
some of the changes can be explained by reasoning drawing on the OTM/ITM
characteristics of the product some of the changes are due to different dependences on
the changed dynamics.
This also highlights the importance of analyzing a range of products with varying
characteristics when comparing models. A failure to do so could result in falsely
concluding that the models did not produce different prices while it would for a
changed characteristic. In the end it would be a matter of which products to price.

While we value the products in our pricing model, the price of the product will
ultimately be determined by the hedging cost of the issuer. In this context the
sensitivity analysis becomes crucial. And we have seen that also for this purpose the
models differ. A wrong hedging parameter would for the issuer wanting to hedge the
exposure in the product result in an imperfect hedge leaving risk on the book of the
issuer,\textsuperscript{61} and for the buyer a wrong picture of the risk in the product.

We know that the models DD, SV and DDSV produce prices of standard caps that are
more close to their market values than the DET model, as the penalty function in the

\textsuperscript{61} As this risk is unwanted and unknown it is the worst kind of risk
calibration is lower for any of these three models. But we have no way of knowing if the prices of the Sticky and Ratchet caps from the extended models are more close to the fair value than the price from the Deterministic model. And more important we do not know if the extended models experience any bias linked to the characteristics of the product, e.g. a mispricing of the product if the margin is too low, stemming from and inadequate modeling of the terminal correlation, or the implied volatility, e.g. mispricings for certain shapes of the volatility surface.

It could therefore be interesting to compare the model prices with market prices of Ratchet and Sticky caps, if these could be obtained. And also repeat the model comparison across different trading days. In this analysis it would be interesting to investigate both the model comparison as performed above, but also the evolution of the model parameters. A symptom of a model that does not capture the dynamics of the forward rates would be fluctuating parameters.

It would also be of great interest to explore further the issue of volatility hedging in the scope of the extended model.
13 Concluding remarks

13.1 Conclusion

The comparison of the models has showed us that the Displaced Diffusion, Stochastic Volatility LIBOR market model of Joshi and Rebonato is theoretically a straightforward extension of the standard model, when assuming the volatility parameterization of this paper. The standard model produces a flat volatility across strikes. The extended model adds a skew by modeling the forward rates as a displaced diffusion. The displaced diffusion was chosen over the constant elasticity of variance model since it induces almost the same skew, with superior modeling properties. The extended model allows a smile in volatility induced by a stochastic volatility. The model maintains the functional form for the instantaneous volatility assumed in the standard model, but lets the four coefficients of the function to be stochastic Ornstein Uhlenbeck processes.

The extended model consequently encompasses the standard model as a special case.

The theoretical analysis of the models also showed us that the interpretation of Black implied volatility as the root mean squared volatility of the forward rate that was possible when the standard model is assumed for the dynamics of the forward rates is no longer possible when assuming the extended dynamics. The Black implied volatility will just be a representation of the price, as it is the number that gives us the right price when put into the Black formula.

In the implementation of the models the main difference is the call for repeated samplings of the volatility in the extended model due to the stochastic volatility. The repeated sampling is necessary in both calibration and pricing. This means that while the standard model can be calibrated within seconds using analytical expressions, the extended model must use numerical methods in the computation of the penalty function. For our implementation this results in a calibration time of the extended model of 2-8 hours depending on the starting parameters. Also when using the models for pricing the computation time is increased considerably.
Both the calibration and pricing time of the extended model means that it cannot be used in a trading environment without improving considerably on the efficiency of the implementation.

In order to analyze the individual effects of the displaced diffusion and the stochastic volatility we calibrated four different models which was subsequently used for pricing: A deterministic, DET, model corresponding to the standard LIBOR market model without skew and smile, a displaced diffusion, DD, model with a skew but no smile, a stochastic volatility, SV, model with a smile but no skew and the full, DDSV, model allowing both a skew and a smile.

When calibrating the models we saw that the extended models allowed a significantly better fit to the market implied volatility surface. We saw that while the two effects, the skew and the smile, each allowed a better fit, combining the two resulted in an improvement in the fit larger than the sum of the improvements of the two separate effects.

Our analysis of prices showed that the models resulted in very different prices. The pricing differences depend on the calibration sample chosen, especially for the DET model, which can only produce a flat volatility surface. We chose to calibrate the DET model to the same volatility surface as the other models, thereby making the parameters of this model a representation of the average volatility across strikes.

We saw that the pricing differences when going from the DET to the extended models differed in both sign and magnitude when the characteristics of the products were changed. This means that the procedure of adjusting qualitatively the prices obtained from a deterministic model to reflect the mispricings compared to the extended model would be very difficult. Also we saw that the effects induced by the skew and the smile were often acting in opposing directions.

While it would be possible to find for each individual product being priced the calibration sample in terms of a cap rate, that would produce a deterministic volatility that would result in the same price as the extended model, the analysis shows that the deterministic volatility would be distinct to that particular product and point in time, and it would be nothing more than a sophisticated Black volatility.
Additionally we observed that the models resulted in differences in hedging parameters, which was examined by calculating sensitivities to changes in the yield curve and the correlation surface.

Recapitulating we have seen that the while the displaced diffusion, stochastic volatility LIBOR market model is a straightforward extension of the standard model, the model results in great differences with respect to implementation, calibration fit and pricing.

13.2 Appraisal of the extended model

The need for repeated samplings of the volatility in both the calibration and the pricing results in very large calibration and computation times. For the calibration this is a great problem aggravated by multiple local minima in the penalty function, which necessitates multiple calibrations to ensure confidence in the result.

Although we have not had focus on efficiency in our implementation, the computation times we report shows that the tractability of the standard model allowing fast calibration and pricing without advanced numerical sampling techniques is lost when moving to the extended model.

The attractiveness of the model therefore depends on the possibility of reducing the computation time by using an implementation that samples the volatility more efficiently. Joshi & Rebonato (2003) report reasonable convergence for 64 volatility paths which would improve the usefulness of the model considerably. But both calibration and pricing would still be at a much higher computational cost than the standard model. Depending on the implementation this could rule out the use of the model in a trading environment. Still we have shown that other relevant uses of the model remains.

The analysis showed that the models resulted in potentially great differences in both prices and hedging parameters, but since it was not possible to evaluate if the prices from the extended model was closer to the fair value of the products, it must be a subject of further research to weight the increase in computation time against the changes in accuracy.
13.3 Suggestions for further work

During the analysis in this paper a number of issues was encountered, which could be subjects for further work.

First of all for a practical implementation it would be valuable to improve on the accuracy and the computation time. The accuracy could be improved by improving the data of the models. We have employed methods characterized by their simplicity and intuition, which suffice for the purpose of our analysis. A practical implementation should therefore address with greater care the issues of the input yield curve, the correlation estimation and last but certainly not least the bootstrap of implied caplet volatilities from the quoted cap volatilities.

As the computation time has been a great concern this issue could also be addressed further in order to make the model more practically applicable. As the increase in computation time as compared to the standard model is a result of the need for repeated sampling of the stochastic volatility, the most successful path to examine would be a more efficient sampling of the volatility enabling fewer volatility paths without deteriorating the accuracy.

Alternatively one could examine the benefits from including the stochastic volatility. The DD model would not require repeated sampling of the volatility and would therefore not be burdened by the same computation time. A weighting of the benefits of going from the DD to the DDSV model against the increase in computation time could therefore be interesting.

As the stochastic volatility requires a different approach to the problem of volatility hedging, this issue could also be an interesting subject of research.
References

Alexander, C., 2002: *Common Correlation Structures for Calibrating the LIBOR Model*, ISMA Discussion paper 18, ISMA Centre, University of Reading


British-Russian Offshore Development Agency: Homepage with Sobol Sequence generator in 1024 dimensions, [www.broda.co.uk](http://www.broda.co.uk)


Verlag


Heston, S., 1993: *A Closed Form Solution for Options with Stochastic Volatility with Application to Bond and Currency Options*, The Review of Financial Studies, vol. 6, no. 2, 327-343

Hull, J., 2003: *Options, Futures and Other Derivatives*, Prentice Hall


Hunter, C.J., P. Jäckel & M.S. Joshi, 2001: *Drift Approximations in the Forward-Rate-Based LIBOR Market Model*, [www.quarchome.org](http://www.quarchome.org), Also published as *Getting the drift*, in Risk, 14, 81-84


Marris, D., 1999: *Financial option pricing and skewed volatility*, MPhil Thesis Statistical Laboratory, University of Cambridge

Moro, B., 1995: *The Full Monte*, Risk, 8, 2

Neftci, S., 2000: *An Introduction to the Mathematics of Financial Derivatives,*
Academic Press


Nicolato, E., 2005: Teaching Materials for the course Financial Engineering at Handelshøjskolen, Aarhus Universitet


Rebonato, R., 1999: *Volatility and Correlation – In the Pricing of Equity, FX and Interest-Rate Options*, John Wiley & Sons


Appendices

Appendix 1: No arbitrage drifts
Appendix 2: Evolving an Ornstein Uhlenbeck process
Appendix 3: The volatility integral in the standard model
Appendix 4: Cholesky decomposition
Appendix 5: Displaced diffusion extension of the Black formula
Appendix 6: Preliminary parameters used in the convergence analysis
Appendix 7: Shifted yield curve and correlation surface

Additional appendices included on CD-ROM
Additional description of the spreadsheets and VBA procedures is included in each workbook

A - Yield curve.xls
   Yield curve data and calibration
B - Correlations.xls
   Correlation data and calibration
C - Volatility.xls
   Volatility data and calibration
D - Pricing Ratchet.xls
   Pricing of Ratchet caps
E - Pricing Sticky.xls
   Pricing of Sticky caps
F - Pricing Standard.xls
   Pricing of Standard caps
G - Hybrid quasi random numbers.xls
   Code to produce hybrid quasi random numbers

The VBA procedures of Pricing Ratchet.xls, Pricing Sticky.xls and Pricing Standard.xls differ only
in the payoff specification
Appendix 1: No arbitrage drifts

The derivation is based on Rebonato (2002).

Consider a set of spanning forward rates $f_i(t)$ resetting at $T_i$, and the associated zero-coupon bonds $P(t,T_j)$ maturing at time $T_j$ in the following denoted by $P_j(t)$.

Also define the natural payoff, $N_i(t)$, of a given forward rate, $f_i$, as that portfolio of traded assets that when multiplied by the forward rate constitutes a traded asset. From the definition of a forward rate it is easy to see that the bond maturing at the same time as the forward rate has its payoff is the natural payoff of the forward rate, i.e.

$$N_i(t) = P_{i^*i}(t)$$

We choose an arbitrary numeraire, $N_j(t)$ from the set of bonds, and define

$$X_N(t) \equiv \frac{N_i(t)}{N_j(t)}$$

and

$$a \equiv f_i(t) X_N(t)$$

We note that both $X_N(t)$ and $a$ are relative prices. Using the results that in order to prevent arbitrage all relative prices must be martingales under the equivalent martingale measure and that all martingales can be written as diffusions we can write the following

$$\frac{df_i}{f_i} = \mu_i dt + \sigma_i dz_i$$

$$\frac{da}{a} = \sigma_a dz_a$$

$$\frac{dX_N(t)}{X_N(t)} = \sigma_{X_N} dz_{X_N}$$
By inserting definitions we get the following for $da$:

$$da = d(f_t X_N)$$

$$= d(f_t) X_N + f_t d(X_N) + f_t \sigma_t X_N \sigma_{X_N} \rho_{f_t X_N} dt$$

$$= \left( f_t \mu_t dt + f_t \sigma_t dz_t \right) X_N + f_t \left( X_N \sigma_{X_N} dX_{X_N} \right) + f_t \sigma_t X_N \sigma_{X_N} \rho_{f_t X_N} dt$$

$$= f_t X_N \mu_t dt + f_t X_N \sigma_t dz_t + f_t X_N \sigma_{X_N} dX_{X_N} + f_t \sigma_t X_N \sigma_{X_N} \rho_{f_t X_N} dt$$

$$= f_t X_N \left( \mu_t + \sigma_t \sigma_{X_N} \rho_{f_t X_N} \right) dt + f_t X_N \sigma_t dz_t + f_t X_N \sigma_{X_N} dX_{X_N}$$

$$\downarrow$$

$$\frac{da}{a} = \frac{d(f_t X_N)}{f_t X_N}$$

$$= \left( \mu_t + \sigma_t \sigma_{X_N} \rho_{f_t X_N} \right) dt + \sigma_t dz_t + \sigma_{X_N} dX_{X_N}$$

$$\downarrow$$

$$\mu_a = \mu_t + \sigma_t \sigma_{X_N} \rho_{f_t X_N}$$

$$\downarrow$$

$$\mu_t = -\sigma_t \sigma_{X_N} \rho_{f_t X_N}$$

The last implication follows because $a$ is a martingale and therefore $\mu_a = 0$

As $\sigma_{X_N}$ is not observable in the market we want to manipulate it further.

For this purpose we define the Vailiant brackets by$^1$

$$\langle x, y \rangle_t = \sigma_x(t) \sigma_y(t) \rho_{x,y} \rho_{x,y}(t)$$

The Vailiant brackets can be interpreted as the instantaneous covariance. The properties of the Vailiant brackets is

$$\langle x, y \rangle_t = \langle x, y \rangle_t + \langle x, z \rangle_t$$

$$\langle x, y \rangle_t = -\langle x, 1/y \rangle_t$$

$^1$ Rebonato modern pricing
Inserting definitions we get

\[ \langle f_i, X_N \rangle = \left( f_i, \frac{N_i(t)}{N_j(t)} \right) = \left( f_i, \frac{P_{ii}(t)}{P_{jj}(t)} \right) \]

From the identity

\[ P_i = \frac{1}{\prod_{x=0}^{x-1} 1 + f_x \tau_x} \]

we can write

\[ \langle f_i, X_N \rangle = \left( f_i, \frac{N_i(t)}{N_j(t)} \right) = \left( f_i, \frac{P_{ii}(t)}{P_{jj}(t)} \right) = \left( f_i, \prod_{k=i+1}^j (1 + f_k \tau_k) \right) = \sum_{k=i+1}^j \langle f_i, 1 + f_k \tau_k \rangle \quad \text{for } j > i \]

\[ \langle f_i, X_N \rangle = \left( f_i, \frac{N_i(t)}{N_j(t)} \right) = \left( f_i, \frac{P_{ii}(t)}{P_{jj}(t)} \right) = \left( f_i, \prod_{k=i+1}^j (1 + f_k \tau_k)^2 \right) = -\sum_{k=i+1}^j \langle f_i, 1 + f_k \tau_k \rangle \quad \text{for } j < i \]

For \( j > i \):\(^2\)

\[ \langle f_i, X_N \rangle = \sum_{k=i+1}^j \langle f_i, 1 + f_k \tau_k \rangle = \sum_{k=i+1}^j \sigma_i \sigma_{i+j} f_i f_j \rho_{f_{i+1} f_{j+1}} \]

From the properties of correlation we have

\[ \rho_{f_{i+1}, f_{j+1}} = \rho_{f_i, f_k} \]

\(^2\) The same derivation can be carried out for \( j < i \)
And a straightforward application of Ito’s lemma gives $\sigma_{1+\tau_{f}}$

\[
df = \mu_{f} dt + \sigma_{f} dz \\
d(1 + f\tau) = (\cdots) dt + \tau \sigma_{f} dz \\
\downarrow \\
\frac{d(1 + f\tau)}{1 + f\tau} = (\cdots) dt + \frac{\tau \sigma_{f}}{1 + f\tau} dz \\
\downarrow \\
\sigma_{(1+f)\tau} = \frac{\tau \sigma_{f}}{1 + f\tau}
\]

Combining this we get

\[
\langle f_{i}, X_{N} \rangle = \sum_{k=i+1}^{j} \sigma_{f_{k}} \frac{\tau \sigma_{f}}{1 + f\tau} \rho_{f_{j}, f_{k}} \\
= \sigma_{f_{i}} \sum_{k=i+1}^{j} \frac{\sigma_{i} \rho_{f_{j}, f_{k}} f\tau}{1 + f\tau}
\]

With full notation this amounts to these drift terms when choosing $P_{j+1}(t)$ as numeraire:

\[
\mu_{i} = \sigma_{i} (t) \sum_{k=j+1}^{i} \frac{\sigma_{k}(t) \rho_{\Delta}(t) f_{k}(t) r_{k}}{1 + f_{k}(t) r_{k}} \\
\text{for } i > j \\
\mu_{i} = -\sigma_{i} (t) \sum_{k=i+1}^{j} \frac{\sigma_{k}(t) \rho_{\Delta}(t) f_{k}(t) r_{k}}{1 + f_{k}(t) r_{k}} \\
\text{for } i < j \\
\mu_{i} = 0 \\
\text{for } i = j
\]
Appendix 2: Evolving an Ornstein Uhlenbeck process

A variable following an Ornstein Uhlenbeck process given by

\[ dY = RS(RL - Y)dt + \sigma dz \]

can be evolved to a future point in time from the current value, \( Y(t) \), by the following exact formula: \(^1\)

\[ Y(T) = RL + (Y(t) - RL)e^{-RT/s} + \frac{\sigma^2}{2R} \left( 1 - e^{-2RT/s} \right) Z \]

where

\( Z \) is a normal random variable

---

\(^1\) Rebonato (2004)
Appendix 3: The volatility integral in the standard model

When the volatility is given by

$$\sigma_i(t) = \left[a + b(T_i - t)\right]e^{-c(T_i - t)} + d$$

the elements of the TOTC matrix are given by

$$TOTC(i, j) =$$

$$\int \sigma_i(t)\sigma_j(t)\rho_{ij} \, dt =$$

$$\int \left[\left[a + b(T_i - t)\right]e^{-c(T_i - t)} + d\right]\left[\left[a + b(T_j - t)\right]e^{-c(T_j - t)} + d\right] \rho_{ij} \, dt =$$

$$\frac{1}{4c^3} \left(e^{-cT_i} \rho_{ij} \left(e^{2ct} (2c^2 (T_i - t)(T_j - t) + c(T_i + T_j - 2t) + 1)b^2 - 2ce^{ct} (-c(T_i + T_j - 2t) - 1) + 2d(e^{ct} (-cT_j + ct - 1) + e^{ct} (-cT_i + ct - 1))\right) + b + 2c^2 (e^{2ct} a^2 + 2d(e^{c(T_i+t)} + e^{c(T_j+t)})a + 2cd^2 e^{2c(T_i+t)})\right)$$

The integrated squared volatility is given as a special case by

$$TOTC(i, i) =$$

$$\int \sigma_i(t)^2 \, dt =$$

$$\int \left[\left[a + b(T_i - t)\right]e^{-c(T_i - t)} + d\right]^2 \, dt =$$

$$\frac{1}{4c^3} \left(e^{-2cT_i} (e^{2ct} (2c^2 (T_i - t)^2 + 2c(T_i - t) + 1)b^2 - 2ce^{ct} (4de^{cT_i} (-cT_i + ct - 1) + ae^{ct} (-2cT_i + 2cT_j - 1)) + b + 2e^2 (e^{2ct} a^2 + 4de^{c(T_i+t)} a + 2cd^2 e^{2cT_i})\right)$$

The integrals are obtained from http://integrals.wolfram.com
Appendix 4: Cholesky decomposition

When implementing the model we will make use of Cholesky decomposition.\(^1\)

The Cholesky decomposition allows generation of correlated random variables. We consider the need for calculating a vector of random variables \( \Psi \) which are correlated with a covariance matrix given by

\[
E(\Psi \Psi^T) = \Sigma
\]

We introduce a matrix \( A \), which satisfy

\[
AA^T = \Sigma
\]

It is then possible to transform a vector of uncorrelated variables \( Z \) to a vector of variables with the appropriate covariance as

\[
\Psi = AZ
\]

It is easy to show that the variables exhibits the correct covariance as

\[
E(\Psi \Psi^T) = E(AZZ^T A^T) = A E(ZZ^T) A^T = AA^T = \Sigma
\]

The decomposition of the covariance matrix into \( A \) can be obtained in a variety of ways. One method is the Cholesky decomposition, which provides the matrix \( A \) as a lower triangular matrix. The actual computation of the Cholesky decomposition is a technical concern and we therefore refer to the code provided in Wilmott (2001).

\(^1\) Wilmott (2001)
Appendix 5: Displaced diffusion extension of the Black formula

As the displaced diffusion just results in a log normal distribution of $f_t + \alpha$ instead of $f_t$, the caplet value is easily seen to be given by

\[
C_i^{\text{Black}} = P(t, T + \tau) \left( (f(t, T + \tau) + \alpha)N(d_1) - (R_K + \alpha)N(d_2) \right) NP \tau
\]

\[
d_1 = \frac{\ln\left( \frac{(f(t, T + \tau) + \alpha)/R_K + \sigma_{\text{Black}(a)}^2(T - t)/2}{\sigma_{\text{Black}(a)} \sqrt{T - t}} \right)}{\sigma_{\text{Black}(a)} \sqrt{T - t}}
\]

\[
d_2 = d_1 - \sigma_{\text{Black}(a)} \sqrt{T - t}
\]

\[
\sigma_{\text{Black}(a)} = \sqrt{\int_1^T \frac{\sigma_a^2(u)du}{T_i}}
\]

where

- $\tau$ is tenor
- $P(t, T)$ is the time $t$ price of the zero coupon bond maturing at time $T$
- $f(t, T + \tau)$ is the forward rate at time $t$ for the period $T$ to $T + \tau$
- $\alpha$ is the displacement coefficient
- $N(x)$ is the standard cumulative normal distribution
- $R_K$ is the cap rate
- $\sigma_a$ is the instantaneous volatility of the displaced diffusion
Appendix 6: Preliminary parameters used in the convergence analysis

As the convergence analysis is performed before the final volatility parameters of the model are obtained, we use the following set of parameters obtained from a preliminary calibration in the convergence analysis.

The correlation and the yield curve is the same as in the final model.

<table>
<thead>
<tr>
<th>RS</th>
<th>RL</th>
<th>sigma</th>
<th>Initial</th>
<th>Alfa</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>3.988</td>
<td>0</td>
<td>3.988</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>12.478</td>
<td>0</td>
<td>12.478</td>
</tr>
<tr>
<td>ln c</td>
<td>1</td>
<td>-0.669</td>
<td>0</td>
<td>-0.669</td>
</tr>
<tr>
<td>ln d</td>
<td>0.011</td>
<td>1.173</td>
<td>0.275</td>
<td>1.173</td>
</tr>
</tbody>
</table>
Appendix 7: Shifted yield curve and correlation surface

As a basis for the analysis of hedging parameters we use the following shifted yield curve and correlation surface. To facilitate comparison we reproduce the unshifted curves and surface.