Modelling Interest Rate Derivatives

Jochen Theis

jochen.theis@sc.com

21/22 September 2016

The present lecture notes aim to give a brief, practical introduction to the techniques used to manage exotic interest rate derivatives in banks today. The practical aspects of interest rate models are typically of just as much importance as their theoretical properties in these applications. In particular, it is necessary to compute not only the prices of a large portfolio of exotic derivative contracts (typically in the tens of thousands), but also their risk characteristics, commonly known as *Greeks*, with respect to tens — sometimes over a hundred — market parameters quickly enough to be of practical use. This drives the choice and the development of interest rate models in this setting.

We first discuss some examples of interest rate derivatives to set the scene. Then we recall the fundamentals of pricing theory and recall basic, model independent replication techniques, which can be very useful to simplify even quite complex pricing problems.

As the implementation aspects of models play such an important role in practice, we will then take a slightly unusual route and use a relatively simple model, a one-factor Gaussian HJM model, also commonly known as the Vasicek-Hull-White model, to discuss implementation techniques.

Only afterwards will we examine more complex interest rate models and discuss their implementation possibilities, which are typically much more restrictive than for the simple model. After covering single currency models, we discuss multi-currency modelling and how to use it to address the need for multi-curve modelling which arose in the Global Financial Crisis.

Contents

1 Introduction

We begin with some examples to illustrate what types of interest rate derivatives are traded and need to be covered by interest rate models. We then recall the theoretical framework for interest rate models and the Heath-Jarrow-Morton Framework.

1.1 Trade Examples

As an illustration of what the models need to accomplish, here are some common types of exotic interest rate structures which we want to be able to price. We start with variants of vanilla option contracts.

1.1.1 Auto- and Flexi-Caps/Floors

These are modifications of standard interest rate cap or floor contracts designed to make them cheaper for the buyer.

Recall that a standard cap or floor contract is a collection of vanilla options on the payments of a standard floating leg. The buyer pays a premium up front and the option seller pays out the option value at the end of each accrual period; for a cap it is the maximum of zero and the difference between the strike and the reference floating rate, for the floor it is the maximum of zero and the difference between the strike and the reference floating rate. In other words, a cap is a basket of call options on individual floating rates, which are called *caplets*, and the floor is a basket of put options, called *floorlets*.

One way to modify a cap to make it cheaper is to restrict the number of caplets that can be exercised to a maximum. If the option seller pays out each caplet that is in the money up to the maximum number, this is called an *auto-cap*. If the buyer can choose which of the caplets are to be paid out then this is called a *flexi-cap*[1].

Quick Exercise. *For the same number of caplets to be exercised, which is cheaper, an auto-cap or a flexi-cap?*

1.1.2 Trigger Options

Another way to make options cheaper is to make the payout dependent on extra conditions. The most common one used in interest rates is to require a second reference rate to be above or below a given level, typically called a trigger level. For example, one may limit a cap on 3M USD LIBOR to pay out only if the 1Y

[1] Note that while these names are common they are not standardised, so you may find that some institutions use different names for these structures.

USD LIBOR resets above a given *trigger level*. Note that applying a trigger to a fixed rate payment is the same as a digital option on the trigger rate.

Quick Exercise. *Write out a term sheet for a trigger cap, i.e. a full description of the cash flows due.*

1.1.3 Range Accruals

Range accruals are a more "continuous" way to apply triggers to payments. The payment is multiplied by the fraction of days during the accrual period on which some trigger rate resets within a given range. For example, a fixed range accrual might pay out a coupon of $X\%$ times n/N where N is the number of calendar days during the accrual period and n is the number of these days on which the applicable 3M USD LIBOR is between $L\%$ and $U\%$. All kinds of variants exist, for example the payoff may be a floating rate or an option, or one may have a more complex accrual condition.

Quick Exercise. *Show that a fixed range accrual can be decomposed into a (large) basket of digital options.*

1.1.4 Spread Options

Spread options work like caps or floors, except that the payout is an option on a difference between two interest rates. The most common ones are CMS spread options, which pay out on the spread between a long-maturity and a short-maturity swap rate, for example the difference between the 30Y and the 2Y USD swap rate. There is typically one pair of swap rate tenors that is commonly referenced for each currency.

Spread options give exposure to the steepness of the yield curve.

1.1.5 Target Range Accruals

Another trick to make trades cheaper (and so allow larger initial coupons to attract investors) is to limit the total payout that the payer of the exotic leg commits to making. In a target range accrual one party pays LIBOR while the other party pays a coupon rate C_i linked to a reference floating rate, typically capped and floored and possibly leveraged, i.e.

$$C_i = \min\left(c_i, \max\left(f_i, \lambda_i R_i + K_i\right)\right)$$

where R_i is the reference rate, $c_i > f_i \geq 0$ are the cap and floor and λ_i and K_i are constants. When the total coupon rate $\sum_i \alpha_i C_i$, where α_i are the accrual fractions, reaches a fixed target T, then the trade terminates. There are variations under which the last coupon can be capped so that the total payout does not exceed T or the total T is made up upon maturity of the trade if the total coupon falls short.

Quick Exercise. *Assume that the total coupon rate paid is close to the target but the trade has several years to run. What is the payout profile for the party paying the exotic leg? What if the yield curve is very steep, i.e. the forward rates for future periods increase quickly? Can you construct a situation in which the next coupon effectively decides whether the trade terminates or runs for another two years?*

1.1.6 Snowballs and Snowblades

Snowballs are swaps in which one party pays a standard floating leg while the other party pays a sequence of coupons that are linked to one another, e.g.

$$C_i = \min\left(\max(\lambda_i R_i + K_i + \mu_i C_{i-1}, f_i), c_i\right)$$

where R_i is a reference floating rate, K_i, λ_i and μ_i are constants and c_i and f_i are a cap and a floor. Payout can vary drastically and either very quickly grow or fall to zero, which led an imaginative marketer to coin the name.

A variant on the snowball is to add a target range accrual feature, which is commonly called a snowblade.

1.1.7 Bermudan Swaptions

Bermudan swaptions are designed to give the buyer more flexibility than a basic (European style) swaption.

Recall that a swaption is an option to enter a vanilla interest rate swap of a given tenor and fixed rate on a given exercise date. If the option holder enters as the fixed rate payer it is called a *payer swaption*, otherwise a *receiver swaption*.

A Bermudan swaption modifies a European swaption by allowing the option buyer to enter into the underlying swap over its remaining life on several dates, at the beginning of every accrual period.

Quick Exercise. *Payer swaptions and caps are options that protect the buyer against high interest rates, while receiver swaptions and floors protect the buyer against low interest rates. Which is cheaper, a cap/floor or a swaption of the same strike? How does a Bermudan swaption compare?*

1.1.8 General Bermudan Callability

Bermudan swaptions are a simple example of introducing callability features to a simpler structure. This is often a way to either improve the risk profile of a trade or to make it cheaper (if the payer of the exotic leg owns the call right). Typically, callability is granted on, or just before, some or all of the reset dates of the reference rates of the exotic leg, and the trade terminates on the following payment date. Examples are callable inverse floaters or callable range accruals.

1.1.9 Quantos and Inflation-linked Derivatives

Another feature that is often seen in trades is that the payoff is in another currency than the reference rate. This is particularly frequent in Asia, where investors want exposure to USD or EUR interest rates but want to invest and receive coupons in their own currency. Instead of converting coupons at the spot FX rate, the interest rate is directly applied to the notional.

A similar feature arises from inflation-linked derivatives, where coupons are applied to inflation-adjusted notionals, which can be viewed as another currency.

1.1.10 Credit-linked Interest Rate Derivatives

The final type of interest rate derivatives we want to introduce are credit-linked ones. Generally, these are interest rate derivatives with the an additional condition linking the payout to whether a reference credit defaults or not, and any of the previously discussed examples could be modified into a credit-linked version.

The simplest case are *extinguishers*, which terminate without further obligations to either side of the derivative contract on default of the reference credit. These can be used to limit counterparty risk exposures. Simple examples are extinguishing swaps and extinguishing swaptions.

1.2 Valuation Adjustments

Since the Global Financial Crisis the focus of interest rate modelling has been shifting from complex derivatives to more accurate pricing of differences between interest rates, which reflect different credit and liquidity risks and the cost of funding for a bank. Since the impact of these differences depends on a bank's total net exposure to the rates, the calculation has to be performed for large collections of trades and is accounted for as a *valuation adjustment* on this collection of trades rather than as an adjustment to the value of an individual position.

Examples are the Counterparty Valuation Adjustment (CVA), which reflects the net exposure to a particular counterparty, and the Funding Valuation Adjustment (FVA), which represents the cost to the bank of funding the traded positions.

Because a large number of trades have to be valued simultaneously within the same model, which may need to cover a number of currencies and asset classes, the modelling challenges are different from those for exotic derivatives. Models are typically simpler, trading off accuracy against speed of calculation.

1.3 Theoretical Framework

Derivative pricing theory is based on reproducing derivative payoffs via hedging strategies in underlying market instruments. The Fundamental Theorem of Asset Pricing characterises markets according to whether such pricing is possible (no arbitrage) and whether it is unique (completeness). We will recall the theorem now.

The underlying assumption is that there is a probability measure and a filtration $\mathcal{F}$ on some probability space such that the measure gives the probabilities of specific events occurring in the real world (hence "real world measure"), and the filtration specifies the information available to the market at a given time. Some technical restrictions have to be imposed on the trading strategies permitted in the model of the market to prevent pathological cases such as making a profit with vanishingly small risk or doubling strategies. Such trading strategies are called *admissible*. There are various possible conditions which depend on the way a general market is modelled; for our purposes we can neglect these technicalities.

A *numeraire* is a way of accounting in an economy. Assets generate cash flows, but money has a time value and will be invested in some form (no rational person keeps it under the mattress!). The numeraire is the security that any cash is invested in. More precisely, a numeraire is any (admissible) tradeable which has strictly positive value at all times.

An *risk-neutral measure* or *equivalent martingale measure* for a numeraire N is a measure equivalent to the real-world measure such that the numeraire-discounted price process of any tradeable in the market is a martingale.

The *Fundamental Theorem of Asset Pricing* states that for any admissible numeraire

1. A market is arbitrage free if, and only if, there exists an equivalent martingale measure.

2. An arbitrage free market is complete if, and only if, the equivalent martingale measure is unique.

There are some technical conditions around these statements that depend on the precise set-up of a general market. We will describe the relevant ones for the HJM framework below.

In practice, interest rate markets are generally assumed to be arbitrage free and complete, or at least completable by adding a few extra products, for the purpose of modelling interest rate exotics. In particular, models are normally based on diffusion processes without jumps. The reason for this is partly a cautious attitude on the trading side, where a simple model with known limitations that allows an intuitive understanding is often preferred to a more complex model whose behaviour is not well understood. Technically, the main reasons for the choice are

- diffusion-based models are well-understood and have been quite successful at modelling and hedging complex derivatives
- additional risk factors can normally be incorporated through additional diffusion-based risk drivers, such as stochastic volatility
- in an incomplete market one has to price by maximising (or minimising) over all possible martingale pricing measures, leading to a vastly more complex problem
- calibrating jump models is far harder than calibrating diffusion-based models, and parameters cannot easily be estimated by analysing market data (compare Girsanov's theorem for jump processes with the diffusion version)

We therefore assume from now on that a unique equivalent martingale measure exists for any admissible numeraire, i.e. for any trading strategy that has strictly positive value at all times (almost surely). This gives us an abstract pricing formula for any derivative: if V is the price process of a tradeable then the process $\frac{V}{N}$ is a martingale under the equivalent martingale measure for N, and therefore the price of the derivative at any time t is

$$V_t = N_t \mathbb{E}\left[\frac{V}{N} \middle| \mathcal{F}_t \right]$$

where the expectation is under the martingale measure.

The simplest tradeables to consider are fixed cash flows at future points in time, i.e. *discount bonds* or *zero coupon bonds*. Let P_{tT}, or alternatively $P(t, T)$, denote the price at time t of the risk free discount bond expiring at time T. Then given a numeraire N the numeraire-adjusted discount bond price $\frac{P_{tT}}{N_t}$ must be a martingale for every T under the N-martingale measure.

Note that the function $T \mapsto P_{tT}$ is the discount curve at a given time t. This can be equivalently represented by the (instantaneous) forward rates defined by $f_{tT} = -\partial_T \log P_{tT}$. Note that since P_{tT} is a valid numeraire and $\frac{P_{tS}}{P_{tT}}$ is a martingale under the associated martingale measure, the instantaneous forward rate $f_{tT} = -\frac{\partial_S|_{S=T} P_{tS}}{P_{tT}}$ is a martingale under this measure as well. This measure is called the T-forward measure. Note that f_{tT} is generally not a martingale under any other forward measure, and there is generally no forward measure under which all forward rates are simultaneously martingales.

This trick of defining particular measures under which particular interest rates become martingales is a very common feature of interest rate modelling since it can materially simplify modelling problems.

1.4 Risk Free vs Market Interest Rates

Models for general interest rate derivatives have to model the behaviour of the whole yield curve (at least up to the maturity of the derivative) over time. This is already quite a complex modelling problem, so standard interest rate models only consider the dynamics of a single yield curve.

Until the recent financial crisis this was adequate; now the differences (spreads) between the standard discount rates (usually an overnight index) embedded in collateral agreements and standard reference floating rates, e.g. LIBOR, have become significant and more volatile — recall the introduction to interest rate markets in Module 3. These need to be incorporated in the modelling.

However, it is useful to start with simple models and to develop a thorough understanding of their characteristics and implementation techniques. Not only are simple models still in use for some purposes, but the costs and benefits of more complex models as well as implementation choices can be better understood by comparison with simpler models. Also, as we will see, we can model multiple curves in a market with the same techniques as for multi currency markets, which we can construct from simpler models for single curves.

1.5 The Heath-Jarrow-Morton Framework

We quickly recapitulate the basic characteristics of the Heath-Jarrow-Morton framework for modelling the term structure of interest rates. The HJM approach is general in the sense that any reasonable, continuous arbitrage free interest rate model can be described as an HJM model (see [*Baxter*, 1997] or chapter 8 of [*Hunt and Kennedy*, 2004]). We first give a brief description of the general model; for details see e.g. [*Heath et al.*, 1992] or chapter 13 of [*Musiela and Rutkowski*, 1997]. The Gaussian version of the model corresponds to deterministic forward rate volatilities, and we describe how Hull-White models relate to them.

The HJM approach assumes a continuous time economy driven by a d-dimensional Brownian motion W on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Over a trading time interval $[0, T^*]$ it describes the dynamics of the instantaneous forward rates for every time $T \leq T^*$. It is assumed that the economy is arbitrage free and complete[2], so that there exists a unique risk-neutral martingale pricing measure.

[2] In full generality the framework can deal with incomplete markets, but we don't need that here.

To simplify notation we assume from here on that $\mathbb{P}$ is this measure.

1.5.1 HJM Assumptions

More precisely, the HJM approach assumes that

$$f_{tT} = f_{0T} + \int_0^t \alpha_{uT} du + \int_0^t \sigma_{uT} dW_u \tag{1}$$

for all $t \in [0, T]$, where

1. the initial forward rate curve $f_{0\cdot} : [0, T^*] \to \mathbb{R}$ is Borel measurable, and
2. $\alpha : C \times \Omega \to \mathbb{R}$ and $\sigma : C \times \Omega \to L(\mathbb{R}^d, \mathbb{R})$ are functions, where $C = \{(u,t)|0 \le u \le t \le T^*\}$ and $L(V, W)$ denotes the space of linear maps from a vector space V to a vector space W. Furthermore, for every maturity T, $\alpha_{\cdot T}$ and $\sigma_{\cdot T}$ follow adapted processes such that

$$\int_0^T |\alpha_{uT}| du + \int_0^T ||\sigma_{uT}||^2 du < \infty, \quad \mathbb{P}\text{-a.s.}$$

   where $||\sigma_{uT}||^2 = \sigma_{uT}\sigma^*_{uT}$ is the Euclidean norm of σ_{uT} and * denotes the adjoint.

The conditions of arbitrage freeness and completeness in fact completely determine α as

$$\alpha_{tT} = -\sigma_{tT}\Sigma^*_{tT}, \qquad \text{where} \quad \Sigma_{tT} = -\int_t^T \sigma_{tu} du. \tag{2}$$

The condition will result from the calculations we carry out below.

1.5.2 Discount Bond and Numeraire Dynamics

With the above assumptions, the discount bond prices process is

$$\begin{aligned}
P_{tT} &= \exp\left(-\int_t^T f_{ts} ds\right) \\
&= \exp\left(-\int_t^T \left(f_{0s} + \int_0^t \sigma_{us} dW_u + \int_0^t \alpha_{us} du\right) ds\right) \\
&= \frac{P_{0T}}{P_{0t}} \exp\left(-\int_0^t \left(\int_t^T \sigma_{us} ds\right) dW_u - \int_0^t \left(\int_t^T \alpha_{us} ds\right) du\right) \\
&= \frac{P_{0T}}{P_{0t}} \exp\left(\int_0^t (\Sigma_{uT} - \Sigma_{ut}) dW_u - \int_0^t (A_{uT} - A_{ut}) du\right) \qquad (3)
\end{aligned}$$

where $A_{tT} = \int_t^T \alpha_{ts} ds$. The money market account B_t, which is the numeraire, evolves as

$$
\begin{aligned}
B_t &= \exp\left(\int_0^t f_{ss} ds\right) \\
&= \exp\left(\int_0^t \left(f_{0s} + \int_0^s \sigma_{us} dW_u + \int_0^s \alpha_{us} du\right) ds\right) \\
&= \frac{1}{P_{0t}} \exp\left(\int_0^t \left(\int_u^t \sigma_{us} ds\right) dW_u + \int_0^t \left(\int_u^t \alpha_{us} ds\right) du\right) \\
&= \frac{1}{P_{0t}} \exp\left(-\int_0^t \Sigma_{ut} dW_u + \int_0^t A_{ut} du\right). \qquad (4)
\end{aligned}
$$

Consequently, the numeraire-adjusted discount bond price $\frac{P_{tT}}{B_t}$ is

$$\frac{P_{tT}}{B_t} = P_{0T} \exp\left(\int_0^t \Sigma_{uT} dW_u - \int_0^t A_{uT} du\right) \qquad (5)$$

As P_{tT} is tradeable, $\frac{P_{tT}}{B_t}$ must be a martingale under the risk-neutral measure $\mathbb{P}$, so we must have

$$\int_0^t A_{uT} du = \frac{1}{2} \int_0^t \Sigma_{uT} \Sigma_{uT}^* du \qquad (6)$$

for every $t \leq T$, hence $A_{tT} = \frac{1}{2} \Sigma_{tT} \Sigma_{tT}^*$ for every $t \leq T$ and so we recover the no-arbitrage condition $\alpha_{tT} = -\sigma_{tT} \Sigma_{tT}^*$.

1.5.3 Forward Measures

For any T, the martingale $M_t = \frac{B_0}{P_{0T}} \frac{P_{tT}}{B_t}$ is the Radon-Nikodym density $\frac{d\mathbb{P}_T}{d\mathbb{P}}$ of the martingale pricing measure $\mathbb{P}_T$ associated to the numeraire $P_{\cdot T}$, the T-forward measure, with respect to the risk-neutral measure $\mathbb{P}$. By Girsanov's theorem, $W^T = W - M^{-1} \bullet \langle M, W \rangle = W - \int_0^{\cdot} \Sigma_{uT}^* du$ is a Brownian motion under $\mathbb{P}_T$, and for all $S \leq T^*$ and $t \leq S \wedge T$

$$\frac{P_{tS}}{B_t} = P_{0S} \exp\left(\int_0^t \Sigma_{uS} dW_u^T + \int_0^t \Sigma_{uS} \left(\Sigma_{uT}^* - \frac{1}{2} \Sigma_{uS}^*\right) du\right) \qquad (7)$$

so that the numeraire-adjusted price process of a discount bond is

$$\frac{P_{tS}}{P_{tT}} = \frac{P_{0S}}{P_{0T}} \exp\left(\int_0^t (\Sigma_{uS} - \Sigma_{uT}) dW_u^T - \frac{1}{2} \int_0^t ||\Sigma_{uS} - \Sigma_{uT}||^2 du\right) \qquad (8)$$

under this measure. Note that the forward measure for the time horizon T^* is commonly called the *terminal measure*.

Exercise 1. *Check the derivation of equations (7) and (8).*

2 Model Independent Pricing: Static Replication

Before we introduce any concrete interest rate models, we first look at how far we can get in pricing interest rate derivatives without any explicit modelling. This is useful for several reasons:

- modelling assumptions introduce *model risk*, i.e. uncertainty about the value computed
- model independent techniques can often be used to reduce the valuation of a derivative to a simpler, less material problem.

We start with a brief recapitulation of how interest markets work.

Interest rate derivatives are contracts to pay cash flows that are defined in terms of market observable interest rates or interest rate instrument prices. As interest rates are mostly determined in interbank markets, continuous price observations are typically not possible, and most instruments will reference official daily rate fixings.

2.1 Forward Rates

The most basic interest rates determined in the interbank market are the inter-bank offer rates, such as LIBORs. These are typically set as an average of deposit rates offered by a group of large banks for periods (tenors) from one week to one year. The exact details differ from market to market; for example, LIBORs (except for GBP) are determined two working days before the deposit period begins (recall the introduction to interest rate markets in Module 2).

Forward Rate Agreements (FRAs) provide a way to trade rate expectations. A FRA is an agreement between two counterparties in which one party agrees to pay LIBOR on a set notional N over a given time interval, from time T_0 to time T_1, say, in exchange for a fixed payment. The fixed payment is quoted as an interest rate K, so that effectively the first counterparty pays $N\alpha(L-K)$ at the end of the accrual period of the LIBOR, where α denotes the accrual fraction for the time interval from T_0 to T_1, and L the LIBOR fixing, which we assume to be fixed at time T. In an arbitrage-free market, there must exist a risk-neutral measure $\mathbb{P}_{T_1}$ associated to the discount bond expiring at T_1, the end of the accrual period. This is commonly referred to as the *time T_1 forward measure*. In this measure, the value of the FRA at time $t \leq T$ is

$$\begin{aligned}
\mathrm{FRA}_t &= P_{tT_1}\mathbb{E}_{\mathbb{P}_{T_1}}\left[\left.\frac{LP_{TT_1} - KP_{TT_1}}{P_{TT_1}}\right|\mathcal{F}_t\right] \\
&= P_{tT_1}\left(\mathbb{E}_{\mathbb{P}_{T_1}}[L|\mathcal{F}_t] - K\right) = P_{tT_1}(F_t - K), \qquad (9)
\end{aligned}$$

where $F_t = \mathbb{E}_{\mathbb{P}_{T_1}}[L|\mathcal{F}_t]$. F_t, which is a martingale under $\mathbb{P}_{T_1}$, is the *forward rate* for the LIBOR at time t.

For longer time periods, for two years and longer, vanilla interest rate swaps are the most liquidly traded instruments in the interbank markets. A vanilla interest rate swap is an agreement to exchange "floating" interest, i.e. interest determined in regular intervals (e.g. every half year) for a fixed rate payment. The floating rate payments are called the "floating leg" of the swap, while the fixed rate payments are called the "fixed leg". A vanilla interest rate swap can be regarded as a portfolio of contiguous FRAs (at least in the USD market where fixed and floating legs have the same frequency), and conversely a FRA can be regarded as a vanilla interest rate swap over a single LIBOR period[3].

Forward swap agreements, like FRAs, can be used to trade swap rates on a forward basis. Like forward LIBORs, forward vanilla swap rates are martingales under their natural martingale measure, which is the martingale measure associated to the numeraire given by the *swap annuity* or *basis point value (BPV)* of the fixed leg of the swap, i.e. $B_t = \sum_{i=1}^n \alpha_i P_{tT_i}$ where the T_i are the payment dates of the fixed leg and the α_i are the accrual fractions of the individual accrual periods.

Exercise 2. *Write out the definition of a vanilla swap rate and prove that it is a martingale in its natural measure.* Hint: follow equation (9).

2.2 Vanilla Options and European Payoffs

Options on forward LIBOR and swap rates are typically also liquidly traded, albeit not necessarily for all strikes. It is normal market convention to quote their prices as implied volatilities in the Black model. The Black model assumes that the forward rates follow a log-normal process with constant volatility, i.e. the SDE of the forward rate is assumed to be

$$dF_t = F_t \sigma_t dW_t$$

where W is a Brownian motion under the martingale measure for the forward rate, so that

$$F_t = F_0 \exp\left(\int_0^t \sigma_s dW_s - \frac{1}{2}\int_0^t \sigma_s^2 ds\right).$$

The price at time t of a caplet, i.e. a call option on a forward rate resetting at T and accruing from T_0 and to T_1, with strike $K > 0$, is

$$\mathrm{Caplet}(F_t, K, \sigma, T-t) = P_{tT_1}\mathbb{E}\left[\max(F_T - K, 0)|\mathcal{F}_t\right] = P_{tT_1}\left(F_t \mathcal{N}(d_+) - K\mathcal{N}(d_-)\right) \tag{10}$$

[3]The frequency of fixed leg payments does not generally match the frequency of floating leg payments, and there are also subtle timing differences between standard swaps and FRAs

where

$$d_{\pm} = \frac{\log \frac{F_t}{K} \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}} \tag{11}$$

and $\mathcal{N}$ denotes the cumulative normal distribution $\mathcal{N}(x) = \int_{-\infty}^{x} e^{-\frac{\xi^2}{2}} \frac{d\xi}{\sqrt{2\pi}}$.

Exercise 3. *Prove equation (10).*

2.3 Static Replication of European Payoffs

While the Black model is used as a convention for quoting, it is *not* an exact model of the market, as market quoted implied volatilities are not constant across strikes, which the model would imply. However, this only implies that delta-hedging with forward rate agreements according to the Black model is not a good way to replicate derivatives; as the vanilla options are liquidly traded, one can use these to hedge.

This is straightforward for European payoffs, i.e. derivatives which pay an amount depending only on the final (reset) value of the rate in units of its natural numeraire. That is, a payoff referencing a LIBOR L accruing from time T_0 to time T_1 that pays a function of L, $f(L)$, at T_1, or a payoff referencing a vanilla swap rate $f(S)$ that pays a function of S in units of the swap numeraire.

If we denote the numeraire by N, the forward rate by F and the reset time by T, then the value of the payoff at t ($\leq T$) is

$$V_t = N_t \mathbb{E}_N \left[f(F_T) | \mathcal{F}_t \right], \tag{12}$$

where $\mathbb{E}_N$ denotes the expectation with respect to the martingale measure associated to the numeraire. The price V_t hence only depends on the distribution of F_T under this measure. If this has a density ϕ_F we can write this as

$$V_t = N_t \int_{-\infty}^{\infty} f(x)\phi_F(x)dx \tag{13}$$

Specialising f to be $f(x) = \max(x-K, 0)$, we obtain from (13) for the call option with strike K

$$\mathrm{Call}(K;t) = N_t \mathbb{E}_N \left[\max(F_T - K, 0) | \mathcal{F}_t\right] = N_t \int_{-\infty}^{\infty} \max(x-K, 0)\phi_F(x)dx. \tag{14}$$

Differentiating twice with respect to K we obtain

$$\frac{\partial^2}{\partial K^2} \frac{\mathrm{Call}(K;0)}{N_0} = \frac{\partial}{\partial K}\left(-\int_K^{\infty} \phi_F(x)dx\right) = \phi_F(K). \tag{15}$$

Hence we obtain for the price of the general European payoff

$$V_t = \int_{-\infty}^{\infty} f(K) \frac{\partial^2 \operatorname{Call}(K;0)}{\partial K^2} dK \tag{16}$$

Note that since call option prices for only a few strikes are observable in practice, it is the assumptions made in the construction of the implied volatility surface that determine the smoothness of the call option price function. A construction method that does not ensure differentiability to at least second order is not advisable.

If $f(K)\frac{\partial \operatorname{Call}(K;0)}{\partial K} \to 0$ and $\frac{\partial f(K)}{\partial K} \operatorname{Call}(K;0) \to 0$ for $K \to \pm\infty$, then the boundary terms vanish when we integrate partially twice and we get

$$V_t = \int_{-\infty}^{\infty} \frac{\partial^2 f(K)}{\partial K^2} \operatorname{Call}(K;0) dK \tag{17}$$

Note that since the call price as a function of strike and its derivatives can normally be expected to converge to 0 for $K \to \infty$ and the support of the density ϕ_F is normally bounded below, the limit conditions are usually satisfied.

Equation (17) extends to cases where the function f is not twice continuously differentiable if we interpret the second derivative of f as a generalised function (or distribution). The right-hand side gives the hedge in terms of calls: the second derivative of the function f is the density with respect to the strike of the position in calls we need to take to hedge the payoff.

Exercise 4. *Carefully check the derivation of equation (16) and work out the technical conditions that f and the density ϕ_F need to satisfy for the proof to work. Then check that the formula gives the correct price for the special cases of a call option and a digital option, i.e.* $f(x) = \max(x - K, 0)$ *and* $f(x) = \mathbf{1}_{\{x>K\}}$.

2.4 LIBOR in Arrears and CMS

The above analysis can be extended to some European payoffs which are not in units of the numeraire. A simple example is LIBOR in arrears, which is LIBOR paid at the beginning of its accrual period rather than the end, i.e. the floating leg pays LIBOR fixed at the end of each accrual period rather than at the beginning. From the general valuation formula (12), its value is

$$V_t = P_{tT_1} \mathbb{E}_{T_1} \left[\frac{F_T}{P_{T_0 T_1}} \middle| \mathcal{F}_t \right] = P_{tT_1} \mathbb{E}_{T_1} \left[F_T (1 + \alpha F_T) | \mathcal{F}_t \right] \tag{18}$$

fits into the static replication framework above, where we write $\mathbb{E}_{T_1}$ as shorthand for $\mathbb{E}_{\mathbb{P}_{\cdot T_1}}$. Other payment times for LIBOR can be treated similarly if one approximates the rate determining the discounting relative to the LIBOR numeraire by the LIBOR itself.

Another common variation are one-time cash payments based on a swap rate, typically made at the beginning or the end of the accrual period of a LIBOR resetting at the same time as the swap rate. These are commonly termed *Constant Maturity Swap* or *CMS* payments. One way to price them is to use the approximation that interest rates are flat at the level of the swap rate over the period of the swap to value the swap numeraire. Another possibility is to use *cash-settled* swaptions rather than physically settled ones for the static hedge. In fact, cash-settled swaptions are as liquid as physically settled ones, and they pay out a cash amount on expiry calculated using the same flat yield curve approximation described above.

Exercise 5. *Write out the valuation formula for a CMS forward payment and for a CMS call option using the flat yield curve approximation. Compute valuation formulas in the Black model.*

Exercise 6. *Show that cash-settled swaptions give the density of a swap rate in the forward measure associated to the payment time of the swaption.*

2.5 The Limits of Static Replication

While European payoffs on a single rate can usually be handled with the static replication methods described above, the method quickly reaches its limits when more than one rate or more than one (reset) time horizon are involved.

When a payoff depends on two rates resetting at the same time, for example an option on the spread between two CMS rates, then it is typically not enough to know just the distributions of the individual rates, but one needs their *joint* distribution. The vanilla call option prices are not enough to give this information, and one needs to make modelling assumptions to compute prices.

Exercise 7. *Consider two rates R and S resetting at the same time and their two-dimensional joint distribution $\phi(x, y) : \mathbb{R}^2 \to \mathbb{R}$. What information on ϕ can you obtain from the vanilla option prices on the individual rates? Can you reconstruct ϕ if, in addition, you also have the prices of all spread options on the rates?*

When the marginal distributions are known, then the possible joint distributions are given by the *copula* functions. A two-dimensional copula is basically a two-dimensional distribution function on $[0, 1] \times [0, 1]$. Given a two-dimensional copula and two distribution functions F and G one obtains a two-dimensional distribution H with marginal distributions F and G by defining

$$H(x, y) = C(F(x), G(y))$$

By *Sklar's Theorem* all two-dimensional distributions with marginals F and G can be constructed in this manner. The simplest example of a two-dimensional copula is the uniform distribution on the unit square $[0, 1] \times [0, 1]$; applying it to given marginals gives the product distribution of the marginals, i.e. the unique joint distribution under which the coordinate random variables are independent with the given marginal distributions.

So one can model the joint distribution of two forward rates, e.g. for a CMS spread option, by *choosing* a copula function to construct a joint distribution. The choice of copula function is a modelling choice and must be justified.

The simplest way to construct two-dimensional copulas is to use a known and easily computable two-dimensional distribution function and to transform it to the unit square so as to make the marginals uniform. In Finance, the most commonly used such copula is the *Gaussian* one: let $\Phi_\rho(x, y)$ be the distribution function of a two-dimensional Gaussian distribution with normal marginals and correlation ρ, i.e. the distribution density is

$$\phi_\rho(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}},$$

and let $\Phi = \int_{-\infty}^{x} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}}$ be the normal distribution function. Then the normal copula function with correlation ρ is given by $C_\rho(x, y) = \Phi_\rho(\Phi^{-1}(x), \Phi^{-1}(y))$. For given marginals F and G the joint distribution given by the Gaussian copula with parameter ρ is then

$$H(x, y) = C(F(x), G(y)) = \Phi_\rho(\Phi^{-1}(F(x)), \Phi^{-1}(G(x)))$$

The Gaussian copula is popular because it gives the natural and well-understood joint distribution of for two Gaussian marginals, which usually underlie the basic models. However, it has the same drawback as Gaussian (or log-normal) distributions for the underlying in that it assumes that correlation between the marginals is constant, which typically underestimates risks in the tail of the distribution.

We will not pursue the topic of copulas further, but similar techniques will be discussed later in the context of Markov Functional Models.

2.6 Semi-static Replication as a Numerical Technique

While static replication is usually no longer possible if payoffs are not European style, the techniques can in some cases be used for effective implementation. We give the simple example of a discrete FX or Equity barrier option here to illustrate this.

The barrier option pays 1 at the payment time T if the spot FX rate (or equity spot price) X_t stays between K_0 and K_1 on dates $t_1 < \cdots < t_n \leq T$ and 0 otherwise.

For the case $n = 1$ we already know how to construct a static hedge. Now assume we have put that hedge in place to cover the last observation time. Consider the observation time immediately prior, t_{n-1}. If $X_{t_{n-1}}$ lies between K_0 and K_1, then we'll be precisely hedged. However, if $X_{t_{n-1}}$ is realised below K_0 or above K_1, then we are over-hedged. If we know the value of the hedge portfolio at t_{n-1} and if we have options on X_t expiring at t_{n-1} available to hedge, then we can put a static replication hedge in place for this over-hedge. If $X_{t_{n-1}}$ comes to lie outside the interval $[K_0, K_1]$, then we will need to sell the hedge portfolio at t_{n-1}. By induction, we can construct hedge portfolios at the prior observation times $t_{n-2}, \dots, t_1$ in the same manner.

It is important to note, that we had to make two assumptions for this. The less critical one is that we have a options for all expiries $t_1, \dots, t_n$ and all strikes available for hedging. The other, more critical assumption was that we know the value of the remaining hedge portfolio (contingent on the realisation of X_t up to that point) at all observation times. Since the hedge portfolio consists, by construction, of vanilla options, this means that we must know the *forward* prices of vanilla options expiring at time t_i for all times $t_j < t_i$. These are generally not traded, even for exchange traded products, and hence it is necessary to use an explicit modelling assumption to compute these.

However, within a given model the forward price of an option is determined as a function of the forward state of the market. The construction above works if the dependency is only on the forward market price X_t. In fact, it works as long as there is a single state variable on which the forward option price depends for which options are available to hedge the exposure.

Exercise 8. *Use the technique described above to compute the price of a discretely observed barrier option in the Black model. For simplicity, you may want to consider only the case of a single barrier.*

What happens when you have more frequent observations? What do you obtain for the limit case of continuous observations? Use the reflection principle to price a continuously observed barrier option directly and compare your results.

As we have seen, as soon as more than one rate or more than a single observation time are involved, we can no longer avoid the choice of a model by hedging with vanilla options. However, given a model, the (semi-)static replication techniques can be very useful to compute the prices, and in particular the hedge portfolios, of particular derivatives. Also, to reduce model risk it is often useful to divide payoffs into a part that can be exactly priced by static replication and a remainder, which will then be priced on a specific model.

3 Vasicek-Hull-White

Having reached the limits of simple pricing techniques by static replication, we now look at a simple classic pricing model, the Generalised Vasicek or Vasicek-Hull-White Model. Rather than rush straight on to more advanced models from there, we will spend some time on implementation methodologies.

We begin by recalling some basic facts about the Hull-White framework and then analyse the generalised Vasicek model.

3.1 The Hull-White Framework

The Hull-White model framework prescribes a mean-reverting diffusion process for the short rate r_t under a martingale probability measure. In full generality

$$dr_t = (a(t) - b(t)r_t)dt + \sigma(t)r_t^\beta dW_t$$

for some constant $\beta \geq 0$, where W is a Brownian motion under the measure. For $\beta = 0$ the short rate has Gaussian law at any time; this is also known as the *generalised Vasicek model* or the *Vasicek-Hull-White model*.

Reminder. *The short rate is an idealisation used in continuous time interest rate models. It represents the instantaneous interest rate in the market and is not directly observable. Relative to the short rate, the prices of discount bonds in the risk-neutral measure are determined by*

$$P_{tT} = \mathbb{E}\left[\exp\left(-\int_t^T r_s ds\right)\middle|\mathcal{F}_t\right]. \tag{19}$$

3.1.1 One-factor (Generalised) Vasicek

The solution to the stochastic differential equation

$$dr_t = (a(t) - b(t)r_t)dt + \sigma(t)dW_t \tag{20}$$

is

$$r_t = e^{-\beta(t)}\left(r_0 + \int_0^t e^{\beta(u)}a(u)du + \int_0^t e^{\beta(u)}\sigma(u)dW_u\right) \tag{21}$$

where $\beta(t) = \int_0^t b(u)du$.

Exercise 9. *Check that (21) solves (20).*

As mentioned above, the price process of the discount bond expiring at T is given by (19) while the money market account is $B_t = \exp\left(\int_0^t r_s ds\right)$. As the underlying measure is the (risk-neutral) martingale probability measure, $\frac{P_{tT}}{B_t}$ must be a martingale for any T. Because

$$\int_0^T r_s ds = \int_0^T e^{-\beta(s)}\left(r_0 + \int_0^s e^{\beta(u)} a(u) du\right) ds \\ + \int_0^T \left(\int_u^T e^{-\beta(s)} ds\right) e^{\beta(u)} \sigma(u) dW_u$$

is a Gaussian process, it follows that

$$\begin{aligned} \frac{P_{tT}}{B_t} &= \mathbb{E}\left[\exp\left(-\int_0^T r_s ds\right)\middle| \mathcal{F}_t\right] \\ &= P_{0T} \exp\left(-\int_0^t (\phi_T - \phi_u) g_u dW_u - \frac{1}{2}\int_0^t (\phi_T - \phi_u)^2 g_u^2 du\right) \end{aligned} \tag{22}$$

where $\phi_t = \int_0^t e^{-\beta(u)} du$ and $g_t = e^{\beta(t)}\sigma(t)$. Hence

$$\begin{aligned} f_{tT} &= -\partial_T \log P_{tT} = -\partial_T \log \frac{P_{tT}}{B_t} \\ &= f_{0T} + \frac{\partial \phi_T}{\partial T}\int_0^t g_u dW_u + \frac{\partial \phi_T}{\partial T}\int_0^t (\phi_T - \phi_u) g_u^2 du \end{aligned} \tag{23}$$

Exercise 10. *Check the derivation of equations (22) and (23).*

This shows, as we will see when we discuss the Heath-Jarrow-Morton framework later, that a generalised Vasicek model is a Gaussian HJM model with separable volatility function $\sigma_{tT} = g_t \frac{\partial \phi_T}{\partial T} = \sigma(t) e^{\beta(t)-\beta(T)}$. The discount bond volatility is consequently $\Sigma_{tT} = g_t(\phi_t - \phi_T)$.

3.2 Forward Measures

The representation of the discount bond dynamics in (22) is not ideal, as we need to know the joint distribution of the two random variables $\int_0^t g_u dW_u$ and $\int_0^t \phi_u g_u dW_u$ to describe all discount bonds. It is clear from the derivation that the second term is linked to the money market account B_t, which is our numeraire. So we can hope to be able to eliminate it by changing numeraire.

For any T, the martingale $M_t = \frac{B_0}{P_{0T}}\frac{P_{tT}}{B_t}$ is the Radon-Nikodym density $\frac{d\mathbb{P}_T}{d\mathbb{P}}$ of the martingale pricing measure $\mathbb{P}_T$ associated to the numeraire $P_{\cdot T}$ with respect to the risk-neutral measure $\mathbb{P}$. $\mathbb{P}_T$ is also commonly called the *time T forward* *measure*. By Girsanov's theorem, $W^T = W - M^{-1} \bullet \langle M, W \rangle = W - \int_0^{\cdot} \Sigma^*_{uT} du$ is a Brownian motion under $\mathbb{P}_T$, and for all $S \leq T^*$ and $t \leq S \wedge T$

$$\frac{P_{tS}}{B_t} = P_{0S} \exp\biggl(\int_0^t g_u(\phi_u - \phi_S) dW_u^T \\ + \int_0^t g_u(\phi_u - \phi_S) \Bigl(g_u(\phi_u - \phi_T) - \frac{1}{2} g_u(\phi_u - \phi_S) \Bigr) du \biggr) \tag{24}$$

so that the numeraire-adjusted price process of a discount bond is

$$\frac{P_{tS}}{P_{tT}} = \frac{P_{0S}}{P_{0T}} \exp\Bigl((\phi_T - \phi_S) \int_0^t g_u dW_u^T - \frac{1}{2} \int_0^t g_u^2 \|\phi_T - \phi_S\|^2 du \Bigr) \tag{25}$$

under this measure. Note that the forward measure for the time horizon T^* is commonly called the *terminal measure*.

Exercise 11. *Check the derivation of equations (24) and (25).*

Equation (25) shows that in the T-forward measure the model is completely described by the stochastic process $\int_0^t g_u dW_u^T$. This is a Markov process and all discount bond prices are functions of this process at any point in time (sometimes called *Markov functionals*).

We can now derive the dynamics of a forward LIBOR under this model: a LIBOR with value date S and maturity date T has forward rate

$$\begin{aligned} F(t) &= \frac{1}{\alpha} \Bigl(\frac{P_{tS}}{P_{tT}} - 1 \Bigr) \\ &= \frac{1}{\alpha} \Bigl(\frac{P_{0S}}{P_{0T}} \exp\Bigl((\phi_T - \phi_S) \int_0^t g_u dW_u^T - \frac{(\phi_T - \phi_S)^2}{2} \int_0^t g_u^2 du \Bigr) - 1 \Bigr) \end{aligned} \tag{26}$$

where α is the accrual fraction. Hence LIBOR forward rates have a shifted log-normal distribution with the shift given by one over the accrual fraction.

3.3 Another Useful Measure

While the model has the very attractive representation (25) in any forward measure, picking a forward measure is not always desirable as discount bonds maturing close to the forward measure date have very low volatility. For some purposes, particularly in the multi-currency context, it is desirable to keep the volatility close to the one in the risk-neutral measure.

To do this, one can define the numeraire (in the risk-neutral measure)

$$N_t = B_t \exp\Bigl(- \int_0^t \phi_u g_u dW_u - \frac{1}{2} \int_0^t \phi_u^2 g_u^2 du \Bigr) \tag{27}$$

Clearly, $\frac{N_t}{B_t}$ is a strictly positive martingale, so N_t is a valid numeraire. The corresponding measure is sometimes called the *pricing measure*.

Setting $D_t = (N_t)^{-1}$, the discount bond dynamics in the pricing measure are

$$D_t P_{tT} = \frac{P_{0T}}{P_{0t}} \exp\left(-\phi_T \int_0^t g_u \dot{W}_u - \frac{1}{2}\phi_T^2 \int_0^t g_u^2 du\right) \tag{28}$$

where $\dot{W}$ is a Brownian motion under the pricing measure. Clearly, the pricing measure also represents the model in terms of functionals of a single Markov process.

4 Implementation Methodologies I

Before moving on to more complex models, we first discuss implementation methodologies for the Hull-White model. Because of its Markov functional property, there are a variety of implementation methods to choose from, and we can contrast the advantages and disadvantages of these without getting too distracted by model complexities.

We will consider pricing problems which depend on observing a finite number of yield curve points $T_1, \ldots, T_n$ at a finite number of times $t_1, \ldots, t_m$, i.e. the model implementation must be able to generate expectations of functions of the forward discount bond prices $P_{t_i T_j}$ for all i, j such that $t_i \leq T_j$ back to a valuation time t_0. As interest rate derivatives generally do not depend on continuous observations, this covers all interest rate derivative contracts normally traded. A continuous observation could also be approximated arbitrarily precisely with a large number of observation times, hence this class of problems is sufficient to cover all practical needs.

As we saw above, for the generalised Vasicek model, we only need to model a single stochastic factor, $\int_0^t g_u dW_u^T$, in the T-forward measure (and analogously for the pricing measure). In fact, we only need the Gaussian variables $X_i = \int_0^{t_i} g_u dW_u^T$. The increments $X_i - X_{i-1} = \int_{t_{i-1}}^{t_i} g_u dW_u^T$ are independent Gaussian random variables with variances $\int_{t_{i-1}}^{t_i} g_u^2 du$, and the numeraire-adjusted discount bond prices can be represented as

$$\frac{P_{t_i T_j}}{P_{t_i T}} = \frac{P_{0T_j}}{P_{0T}} \exp\left(-(\phi_{T_j} - \phi_T)X_i - \frac{1}{2}(\phi_{T_j} - \phi_T)^2 G_{t_i}\right) \tag{29}$$

where $G_t = \int_0^t g_u^2 du$.

4.1 Conditional Expectation Methods

The functional representation of in (29) suggests the following approach: the transition density between the Gaussian variables is the heat kernel

$$k(s,t;x,y) = \phi(x-y,t-s). \tag{30}$$

where

$$\phi(x,t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right) \tag{31}$$

is the fundamental solution of the well-studied heat equation $\partial_t \phi = \frac{1}{2}\partial_x^2 \phi$ on $\mathbb{R}$. In particular, the transition between X_i and X_j for $i < j$ is given by $\phi(x_i - x_j, G_{t_j} - G_{t_i})$.

If we can find a representation of the functionals $f(X_j)$ such that we can easily compute the conditional expectation function

$$\mathbb{E}[f(X_j)|X_i] = \int \phi(X_i - x_j, G_{t_j} - G_{t_i}) f(x_j) dx_j \tag{32}$$

then the pricing of derivatives becomes straightforward.

The general pricing procedure is as follows:

1. Represent numeraire-adjusted value of all potential residual payoffs of the derivative at the final time t_m as a Markov functional $V_{m,i}(X_m)$ and set $k = m$.

2. Compute the Markov functional representation of the potential residual values at the next earlier time t_{k-1} using the transition formula

$$\tilde{V}_{k-1,i}(X_{t_{k-1}}) = \int V_{k,i}(y)\phi(X_{t_{k-1}} - y, G_{t_k} - G_{t_{k-1}})dy$$

3. Combine the $\tilde{V}_{k-1,i}$ according to the term sheet to derive Markov functional representations $V_{k-1,i}$ of the residual values the derivative may take at t_{k-1}.

4. While $k > 1$ go to step 2 replacing k by $k-1$.

At the final step we obtain the desired derivative value at time 0 as $V_0(X_0)$.

Binomial or trinomial trees are examples of this; here the tree is set up to approximate the heat kernel on the tree structure, and the conditional expectation integral (32) reduces to taking the weighted sum over the linked tree nodes.

However, trees are very rough approximations of Brownian motions. Numerical PDE solving schemes provide a better way to calculate propagation under diffusion processes. They work well for quite generic problems and are well worth the additional implementation effort over trees.

As Brownian motions and the heat kernel are very well understood there are alternative methods available for this special case. The common implementation uses a piecewise polynomial representation of the functionals. The reason for this is that piecewise polynomial functions can be analytically propagated by the transition kernel.

This works as follows: for $s < t$ and $n \geq 0$ define

$$\begin{aligned} I_n(a,s,t;x) &= \int y^n \mathbf{1}_{\{y\leq a\}} \phi_1(x-y,t-s)dy \\ &= \int_{-\infty}^{a} y^n e^{-\frac{(y-x)^2}{t-s}} \frac{dy}{\sqrt{2\pi(t-s)}} \end{aligned}$$

Then by a simple partial integration

$$\begin{aligned} I_0(a,s,t;x) =& \Phi_1(a-x,t-s) \\ I_1(a,s,t;x) =& a\Phi_1(a-x,t-s) - (t-s)\phi_1(a-x,t-s) \\ I_n(a,s,t;x) =& aI_{n-1}(a,s,t;x) + (n-1)(t-s)I_{n-2}(a,s,t;x) \\ & - (t-s)a^{n-1}\phi_1(a-x,t-s) \end{aligned}$$

where $\Phi_1(x,t) = \int_{-\infty}^{x} \phi_1(y,t)dy$ is the (suitably scaled) cumulative normal distribution.

This is very efficient to evaluate, evolving only one evaluation of the cumulative normal distribution, one of the exponential function and a number of multiplications and additions proportional to the degree of the polynomial. In fact, for low order polynomials the cumulative normal distribution typically dominates the computation time.

So the algorithm works as follows: the functionals at time t_k are represented as piecewise polynomial functions between a given grid of points $x_{k,\alpha}$, $\alpha = 0,\ldots,m_k$. This can be a fixed grid or a variable one, e.g. a fixed one augmented by points at which the payoff function is not smooth.

So for a given functional F_k one has

$$F_k(X_{t_k}) = \sum_{\alpha=1}^{m_k} F_{k,\alpha}(X_{t_k})\mathbf{1}_{\{x_{k,\alpha-1}<x<x_{k,\alpha}\}}$$

where $F_{k,\alpha}(x) = \sum_{j=0}^{d_{k,\alpha}} a_{k,\alpha,j}x^j$ is a polynomial.

Then the conditional expectation at the grid points of an earlier time t_j is

$$\begin{aligned} F_j(x_{j,\beta}) &= \sum_{\alpha=1}^{m_k} \int_{x_{k,\alpha-1}}^{x_{k,\alpha}} F_{k,\alpha}(y)\phi_1(x_{j,\beta}-x,t_k-t_j)dy \\ &= \sum_{\alpha=1}^{m_k} \sum_{i=0}^{d_{k,\alpha}} a_{k,\alpha,i} \Big(I_i(x_{k,\alpha},t_j,t_k;x_{j,\beta}) - I_i(x_{k,\alpha-1},t_j,t_k;x_{j,\beta}) \Big) \end{aligned}$$

Note that this does not depend on fixed grid points $x_{j,\beta}$ but can be evaluated for arbitrary points. Hence the grid can be augmented when required. The formula is exact, so the conditional expectation can be calculated as precisely as desired.

Having worked out the values of F_j at the grid points $x_{j,i}$, an approximation of F_j is constructed by interpolation, e.g. as a spline.

The justification for this is that as the transition operator is a smoothing operator that respects polynomial degrees (more about this later), this ought to give a good approximation of F_j as long as there are enough grid points. Some implementations allow $-\infty$ and ∞ as grid points, hence including some wing behaviour in the functional.

As in other tree or grid models, path dependency can be represented by state variables.

4.2 Monte Carlo

The above algorithms are all based on the idea of representing conditional expectation functions and using transition densities to propagate them backwards in time. One drawback of these methods is that path dependent payouts, such as TARNs or Snowballs, are not straightforward to compute. One needs to introduce state variables and essentially compute parameter dependent expectation functions to price these.

Another drawback is the cost of representing conditional expectation functions, which typically grows exponentially with the dimension of the driving process. A simple rule of thumb is that the error in each dimension is proportional to n^{-2}, where n is the number of points, so for a d-dimensional process the order of convergence is $n^{-2/d}$.

The alternative is to use Monte Carlo simulation to randomly sample the distribution of the paths, which has order of convergence $n^{-1/2}$. For this we need to sample the Gaussian variables X_i; as their increments are independent Gaussians with known variances, this is straightforward to do.

Pricing with a Monte Carlo simulation is simple in principle. We carry out the following algorithm repeatedly for a set number of interest rate paths and average the result:

1. Initialise the number of numeraires held to the equivalent of any cash in the derivative and propagate from the initial market state to the first reset time.

2. Calculate the cash flow at that time generated by the derivative, convert to the equivalent number of numeraire assets and add them to the amount of numeraires held.

3. If we are not at the last reset time, propagate to the next reset time and go to step 2.

4. At the last reset time, convert the number of numeraires held into their equivalent value at the initial time.

We will go into more detail on Monte Carlo later in the context of Market Models, where the technical aspects are more complex. We will, however, discuss one technical aspect of Monte Carlo now.

4.3 Bermudan Callability and Monte Carlo

While path dependent products are straightforward to price with Monte Carlo, callability features pose a difficulty.

The option holder realises the maximal value of the contract by picking at each exercise time the more valuable alternative of keeping the contract (with the remaining optionality) or exercising. This is straightforward if one can compute the future price of the contract, hence Bermudan optionality is easy to value by backwards population on a tree or finite-difference scheme.

In a Monte Carlo simulation we would have to start a sub-Monte Carlo simulation to compute a price at a future time, which would be prohibitively expensive.

4.3.1 Exercise Strategies

A better way is to consider *exercise strategies*. Assume that we have a Bermudan option X to enter into an underlying contract and that we can compute the *exercise value* $V(t)$ at each exercise time. By decomposing a callable/puttable product into a non-callable underlying and an option to enter into a cancelling contract, we can usually reduce our pricing problem to this form.

The value of the option at time t is then

$$X(t) = \sup_{\tau \geq t \text{ stopping time}} N(t)\mathbb{E}\left[\frac{V(\tau)}{N(\tau)} \middle| \mathcal{F}_t \right]$$

where the supremum is over all *stopping times*, that is over all decision strategies that depend only on information available up to the decision time. The supremum is realised for a particular stopping time τ_0, the *optimal exercise strategy*.

We can price the option by finding the optimal exercise strategy and then using it in the Monte Carlo simulation. As the price with the optimal exercise strategy is a supremum, and a close approximation to the strategy will give a good approximation to the price.

Various different approaches to approximating optimal exercise strategies are possible. A straightforward method that illustrates the method well works as follows:

1. Define a parametric family of exercise strategies.
2. Optimise the parameters on a small family of training paths for the maximum value of the exercise right.
3. Price the derivative with the optimised strategy on a new, larger set of training paths.

The third step is essential to obtain an *unbiased* Monte Carlo estimate of the value of the derivative. If we used the same Monte Carlo paths as for the optimisation of the exercise strategies, then overfitting could bias the result.

Typically, one will choose a parametric family of exercise strategies for each exercise date so that they can be optimised separately in sequence, starting with the last exercise date. A sequence of low-dimensional optimisations is typically considerably more efficient than a single high-dimensional one.

This algorithm gives a guaranteed *lower bound* to the price. While the approach itself is straightforward, the choice of a good parametric family of exercise strategies is hard and requires some understanding of the product.

Example. *Consider a* Bermudan payer swaption *based on the dates $t_0, \ldots, t_m$. The contract gives the holder the right at each reset time t_i, $i = 0, \ldots, m-1$, to enter a payer swap struck at a pre-specified strike K and paying at times $t_{i+1}, \ldots, t_m$, or to retain the Bermudan option for the remaining reset times.*

Some simple, but effective parametric exercise strategies are:

1. *Trigger on the coterminal swap rate s_i, i.e. exercise if s_i resets above a level α_i. The α_i are the parameters to optimise; they will typically decrease to the final time. At the final time we exercise if we are in the money, i.e. we set $\alpha_{m-1} = K$.*
2. *Trigger on the exercise value of the option, i.e. exercise if the PV of the underlying swap is above a level v_i and optimise the v_i. The v_i will generally also decrease, and $v_{m-1} = 0$.*

In the above examples, each parameter affects the exercise decision at a particular time only. Its optimal value will not depend on the parameters for prior times.

These strategies do not need to be optimised globally, which is a very high-dimensional problem, but can be optimised one by one, working backwards along the time steps.

A more complicated exercise strategy parametrisation was found by Peter Jäckel in *Jäckel* [2000] through the analysis of the exercise boundary in a non-recombining tree implementation of a LIBOR market model. The exercise strategy is to exercise at time i if the underlying is in the money and

$$f_i(t) - \left[p_1^i \frac{s_{i+1}(0)}{s_{i+1}(t) + p_2^i} + p_3^i \right] > 0$$

where we again optimise the p_j^i stepwise from the end.

4.3.2 Exercise Strategies by Regression

The disadvantage of the approach described above is that appropriate parameterised families of strategies need to be determined for every payoff and the corresponding families implemented. This is cumbersome to maintain in generic Monte Carlo implementations of models, where one would ideally like to have a generic algorithm that only needs a different configuration, not new code.

One way of doing this is to fit the conditional expectation of the intrinsic value by regression. This method was proposed by *Longstaff and Schwartz* [2001]. It has become the implementation method of choice in the industry today. It is based on the fact that in a Markov model any conditional expectation at a time T can be represented as a function of the stochastic processes that make the model Markov. Under very generic conditions, these functions are square integrable, and therefore they can be approximated by a linear combination of a set of basis functions in these variables.

In terms of conditional expectation functions the exercise rule is very simply: the option holder will exercise if, and only if, the value of the exercised option is greater or equal to the value of the remaining contract after not exercising. The former is commonly called the *intrinsic value*, the latter the *extrinsic value* of the contract at the exercise date.

Let X_t denote the a Markov process so that the value of any derivative at the exercise date T can be written as a function of X_T, and let $I(X_T)$ and $E(X_T)$ be the intrinsic and extrinsic value respectively of the derivative contract at T. The the optimal exercise rule is to exercise if, and only if, $I(X_T) - E(X_T) \geq 0$. To obtain an approximation, it is enough to approximate the function $I - E$. Often the intrinsic value can be calculated explicitly, as in the case of a Bermudan swaption, then only E needs to be approximated.

The algorithm consists of calculating such an approximation by a least-squares regression (in other words, an approximate projection onto the linear subspace spanned by the basis functions in the L^2 space of functions of the Markov variables). The calculation follows similar steps to the parametrised methods described above:

1. Define a set of basis functions for each exercise date.
2. Determine the least squares best fit of a linear combination of basis functions to the conditional expectation of the intrinsic minus the extrinsic value of the derivative at each exercise date starting with the last and working backwards in time, using a set of training paths.
3. Price the derivative using a new set of paths, independent of the training paths.

As before, it is essential to use a new set of paths for the pricing to avoid an overfitting bias, so that one obtains a guaranteed lower bound.

The difficulty, as with the choice of parametric exercise strategies, is in choosing a good set of basis functions. However, it is very easy in a model implementation to let the computer generate sets of polynomials in market observables, and these usually produce good fits.

The success of the algorithm is due to the fact that in diffusion models the conditional expectations functions are very smooth and hence usually well approximated by low-dimensional polynomials in standard market observables, such as LIBOR or swap rates or swap annuities. However, it is not difficult to find payoffs for which this is no longer true, e.g. digital options, found in abundance in range accruals.

A number of tweaks are possible. Since only the behaviour of $I - E$ near zero needs to be approximated closely to obtain a good approximation to the exercise rule, one can refine the sampling of $I - E$ that one uses in the least squares fitting. For example, if the option holder can let the derivative expire without any obligation, as in the case of a Bermudan swaption, the extrinsic value must be non-negative, so any path on which the intrinsic value is negative is one on which one will never exercise. Excluding those paths from the regression will therefore usually result in a better fit of the regression function. Similarly, one can exclude points where $I - E$ is very large, or use importance sampling techniques to concentrate the sampling near the zeros of the function.

4.3.3 Upper Bounds

Exercise strategy approximation gives a lower bound on the price by construction. It does not provide a measure of accuracy. To confidently use the method for pricing, we need to understand its accuracy.

Alternative pricing methods are generally expensive to compute and need to be adapted for each new security type. But it is possible to use Monte Carlo to compute an upper bound.

The basic idea is to replace stopping times with random times. This gives the crude upper bound

$$\sup_{\tau \geq t \text{ random time}} N(t)\mathbb{E}\left[\frac{V(\tau)}{N(\tau)}\middle|\mathcal{F}_t\right] = N(t)\mathbb{E}\left[\max_i \frac{V(t_i)}{N(t_i)}\middle|\mathcal{F}_t\right]$$

While this is too crude to be useful, we can subtract a self-financing hedging portfolio H with zero initial price from our derivative without changing its price. Taking the supremum over random times again, we obtain another upper bound

$$N(t)\mathbb{E}\left[\max_i \frac{V(t_i) - H(t_i)}{N(t_i)}\middle|\mathcal{F}_t\right]$$

Rogers [2002] proves that the infimum over all these upper bounds equals the price of the derivative and that the infimum is realised for a particular portfolio given by the martingale part of the *Snell envelope* of V/N. It follows that it is possible to hedge the derivative even if the counterparty exercises with perfect foresight. The price is thus a *seller's price*.

To compute the optimal portfolio we would need to know the full price process of the derivative, which is what we want to compute, so we need to approximate the optimal hedging portfolio.

We can proceed analogously to the lower bound algorithm:

- Choose a parametric family of portfolios H_α.
- Optimise over the parameters α for a small set of training paths.
- Use a different, larger set of paths to price the derivative with the optimised portfolio H_α.

The choice of a good parametric family of portfolios is hard and requires product understanding.

The algorithm is generally less efficient than the lower bound pricing algorithm because

- The portfolio optimisation is generally global, and hence high-dimensional.
- For the upper bound all the pricing paths need to be fully computed to find the pathwise maximum, whereas for the lower bound we only compute to the exercise time.
- We need to price all hedging portfolio instruments at each point along a pricing path, which is generally more time-consuming than the computation of the exercise strategy.

However, since the hedging portfolio minimises variance, we may need fewer paths in the final pricing step than for the lower bound. The optimisation stage tends to be the more time-consuming, though.

A different algorithm was proposed by Andersen and Broadie *Andersen and Broadie* [2004]. The optimal hedging portfolio is known explicitly as the martingale part of the Snell envelope. The Snell envelope is basically the price process of the *unexercised* derivative. To obtain its martingale part, one has to add back the loss of value at the points where it should be exercised.

A lower bound approximation can be used to calculate this. The advantage of this approach is that it dispenses with the need to construct a sensible parametrised family of hedging portfolios and to optimise. The disadvantage is that one needs a sub-Monte Carlo simulation at each time step to compute the Snell envelope, which makes this computationally *very* expensive. It can still be a sensible check to verify the accuracy of a lower bound method.

Algorithms are still being researched. An example is the proposal by *Belomestny et al.* [2009] to calculate an approximation to the Snell envelope as a stochastic integral, in which the integrand function is calculated by regression. According to the paper, this produces fairly good approximations to the upper bound without requiring the nested simulations of the Anderson-Broadie algorithm. However, our own testing did not show good performance for more general derivatives than the special case of Bermudan swaptions tested in the paper, so we cannot recommend this approach unreservedly.

A very paper proposing an upper bound calculation without requiring sub-Monte Carlo simulations is *Joshi and Tang* [2012].

5 Gaussian Heath-Jarrow-Morton and Hull-White Models

The generalised Vasicek model only allows for limited dynamics of the yield curve. Essentially, the g function described the amount of information entering the market while the ϕ function describes the deformation of the yield curve under the random information. The possible configurations of the yield curve are completely determined by the initial shape of the curve and ϕ.

This is a typical feature of short rate models which satisfy one-dimensional equations in the Hull-White framework. We now discuss how these models can be generalised and what constraints a short rate representation imposes. This will motivate the Market Model approach.

5.1 Short Rate Dynamics

The short rate follows the process

$$r_t = f_{tt} = f_{0t} - \int_0^t \sigma_{ut}\Sigma^*_{ut}du + \int_0^t \sigma_{ut}dW_u. \tag{33}$$

Note that this is not a semimartingale representation of r_t, and generally it is not possible to derive one from the above. However, if f_{0T} and σ_{tT} are differentiable with respect to T, then

$$\begin{aligned}\int_0^t \sigma_{ut}dW_u &= \int_0^t \left(\sigma_{uu} + \int_u^t \frac{\partial\sigma_{us}}{\partial T}ds\right)dW_u \\ &= \int_0^t \sigma_{uu}dW_u + \int_0^t \left(\int_0^s \frac{\partial\sigma_{us}}{\partial T}dW_u\right)ds\end{aligned}$$

and

$$\int_0^t \sigma_{ut}\Sigma^*_{ut}du = \int_0^t \sigma_{uu}\Sigma^*_{uu}du + \int_0^t \int_0^s \frac{\partial(\sigma_{us}\Sigma^*_{us})}{\partial s}duds.$$

Hence the semimartingale representation of the short rate is

$$r_t = r_0 + \int_0^t \gamma_u du + \int_0^t \sigma_{uu}dW_u \tag{34}$$

where

$$\gamma_t = -\sigma_{tt}\Sigma^*_{tt} + \frac{\partial f_{0t}}{\partial t} - \int_0^t \frac{\partial(\sigma_{ut}\Sigma^*_{ut})}{\partial t}du + \int_0^t \frac{\partial\sigma_{ut}}{\partial t}dW_u. \tag{35}$$

5.2 Gaussian HJM Models

The most commonly used HJM models assume that the forward rate volatility σ is deterministic. This implies that the instantaneous forward rates f_{tT}, and hence the spot rates $r_t = f_{tt}$, have Gaussian probability laws, and these models are hence called *Gaussian HJM models*.

By (5) and (8) the numeraire-adjusted discount bond prices are Markov processes in the spot and the forward measures. However, in general, r_t is *not* a Markov process under the spot martingale measure. Requiring the short rate to be Markov places restrictions on the volatility term structure. For non-degenerate forward rate volatility functions the short rate process is Markovian if, and only if, σ is *separable*, i.e. there exist functions $g : [0, T^*] \to L(\mathbb{R}^d, \mathbb{R})$ and $h : [0, T^*] \to \mathbb{R}$ such that $\sigma_{tT} = g(t)h(T)$ for all $t \le T \le T^*$ (see appendix A).

As we saw in the previous section, generalised Vasicek models are Gaussian HJM models with a separable volatility structure. Conversely, one checks easily that a Gaussian HJM model with a separable volatility function can be written as a generalised Vasicek model, so a one-factor Gaussian HJM model that is Markov is a generalised Vasicek model.

Exercise 12. *Check that a one-factor Gaussian HJM model with separable volatility function is a generalised Vasicek model by deriving the SDE for the short rate.*

5.3 Multi-factor (Generalised) Vasicek

The Hull-White model can be extended to more than one factor. The multi-factor version of the generalised Vasicek model prescribes the SDE

$$dr_t = a(t)dt + \sum_{i=1}^{n} dr_t^i$$

for the short rate under the martingale probability measure, where

$$dr_t^i = -b_i(t)r_t^i dt + \sigma_i(t)dW_t^i$$

and the W^i are Brownian motions with correlations $\left\langle W_t^i, W_t^j \right\rangle = \rho_t^{ij}$. Analogously to (21),

$$r_t^i = e^{-\beta_i(t)}\left(r_0^i + \int_0^t e^{\beta_i(u)}\sigma_i(u)dW_u^i\right)$$

where $\beta_i(t) = \int_0^t b_i(s)ds$, and

$$r_t = r_0 + \int_0^t a(u)du + \sum_{i=1}^{n} r_t^i \tag{36}$$

Setting $\phi_t^i = \int_0^t e^{-\beta_i(s)}ds$, we obtain

$$\int_0^T r_s ds = \int_0^T \left(r_0 + \int_0^s a(u)du + \sum_{i=1}^{n} e^{-\beta_i(t)} r_0^i\right) ds$$
$$+ \sum_{i=1}^{k} \int_0^T (\phi_T^i - \phi_u^i)e^{\beta_i(u)}\sigma_i(u)dW_u$$

which is Gaussian, so

$$\frac{P_{tT}}{B_t} = P_{0T} \exp\left(-\sum_{i=1}^{n} \int_0^t (\phi_T^i - \phi_u^i)g_u^i dW_u^i\right.$$
$$\left. -\frac{1}{2}\sum_{i,j=1}^{n} \int_0^t (\phi_T^i - \phi_u^i)(\phi_T^j - \phi_u^j)g_u^i g_u^j \rho_u^{ij} du\right) \tag{37}$$

where $g^i_u = e^{\beta^i(t)}\sigma_i(t)$. Therefore

$$f_{tT} = f_{0T} + \sum_{i=1}^{n} \frac{\partial \phi^i_T}{\partial T} \int_0^t g^i_u dW^i_u + \sum_{i,j=1}^{n} \frac{\partial \phi^i_T}{\partial T} \int_0^t (\phi^j_T - \phi^j_u) g^i_u g^j_u \rho^{ij}_u du. \tag{38}$$

Changing coordinates to n orthogonal Brownian motions $\tilde{W}^i$ such that $dW^i_t = c^{ij}_t d\tilde{W}^j_t$ for some functions $c^{ij} : [0, T^*] \to \mathbb{R}$, we finally have

$$f_{tT} = f_{0T} + \sum_{i,j=1}^{n} \frac{\partial \phi^i_T}{\partial T} \left(\int_0^t g^i_u c^{ij}_u d\tilde{W}^j_u + \int_0^t (\phi^j_T - \phi^j_u) g^i_u g^j_u \rho^{ij}_u du \right)$$

which shows that the multi-factor generalised Vasicek model is a multi-factor Gaussian HJM model.

Note that the HJM forward volatility function is of the form

$$\sigma^{1j}_{tT} = \sum_{i=1}^{n} S(t, T)^{ij},$$

where

$$S(t, T)^{ij} = c^{ij}_t g^i_t \frac{\partial \phi^i_T}{\partial T}$$

is separable, as we expect from appendix A, since the vector $(r^i_t)_{i=1,\dots,n}$ of the component short rates is a Markov process. The short rate process $r_t = \sum_{i=1}^n r^i_t$ on its own, however, is generally not Markovian.

5.4 Parametrisation of Vasicek Models

As we saw above, the parameterisation of the generalised Vasicek model by the functions ϕ and G gave a very useful description of the model. We can interpret the function G as the model time, while the ϕ function gives the sensitivity of the individual points of the yield curve to the driving Brownian motion. The special case $G_t = t$ is called the corresponds to the special form of the generalised Vasicek model where the general stochastic differential equation for the short rate (20) is specialised to

$$dr_t = \left(a(t) + \frac{\dot{\sigma}(t)}{\sigma(t)} \right) dt + \sigma(t) dW_t$$

In the HJM framework, this corresponds to a forward volatility function of the form $\sigma_{tT} \equiv \sigma(T)$.

One can similarly parametrise the two-factor generalised Vasicek model. In fact, one derives for the n-factor generalised Vasicek model, analogously to the above,

$$\frac{P_{tS}}{P_{tT}} = \frac{P_{0S}}{P_{0T}} \exp\left(- \sum_{i=1}^{n} (\phi_S^i - \phi_T^i) \int_0^t g_u^i dW_u^{T,i} \right.$$
$$\left. - \frac{1}{2} \sum_{i,j=1}^{n} (\phi_S^i - \phi_T^i)(\phi_S^j - \phi_T^j) \int_0^t g_u^i g_u^j \rho_u^{ij} du \right) \quad (39)$$

in the forward martingale measure $\mathbb{P}^T$. Defining $G_t^{ij} = \int_0^t g_u^i g_u^j \rho_u^{ij} du$ and $G_t^i = G_t^{ii}$, this can be written as

$$\frac{P_{tS}}{P_{tT}} = \frac{P_{0S}}{P_{0T}} \exp\left(- \sum_{i=1}^{n} (\phi_S^i - \phi_T^i) \tilde{W}_{G_t^i}^{T,i} - \frac{1}{2} \sum_{i,j=1}^{n} (\phi_S^i - \phi_T^i)(\phi_S^j - \phi_T^j) G_t^{ij} \right) \quad (40)$$

where $\tilde{W}^{T,i}$ are again time-changed versions of the Brownian motions $W^{T,i}$.

5.5 Implementation Implications

Multi-factor Vasicek models can be implemented with the same methods as the single factor generalised Vasicek model. The conditional expectation methods will be more efficient in dimensions 2 and 3, but for higher-dimensional models Monte Carlo will generally be the more efficient implementation method.

In a generic Heath-Jarrow-Morton model there will not be any Markov variables. As trees, PDE and Markov functional methods, which compute conditional expectation functions, all only work for Markov processes, this only leaves Monte Carlo simulation as a generic implementation method. As the drift function $-\sigma_{tT}\Sigma_{tT}^*$ is generally a state-dependent random process, a generic simulation will need to use short time steps, and hence need a large number of steps, and also present a very high-dimensional Monte Carlo problem.

It is therefore highly advisable to specialise implementations to particular classes of HJM models. Gaussian HJM models, with their deterministic volatility functions, constitute a particularly tractable class of models, and we now describe the possible implementation methods for all the models described above.

5.5.1 Generic Methods for Gaussian HJM Models

We will again consider pricing problems which depend on observing a finite number of yield curve points $T_1, \ldots, T_n$ at a finite number of times $t_1, \ldots, t_k$, i.e. the model implementation must be able to generate expectations of functions of the forward discount bond prices $P_{t_i T_j}$ for all i, j such that $t_i \leq T_j$ back to a valuation time t_0.

As we noted in section 5.2, the numeraire-adjusted discount bond prices are Markov processes in both the risk-neutral and the forward measures. From equations (5) and (8) they are in fact exponential Brownian motions with deterministic volatility functions in any of these measures, so the covariances of the Brownian motions between the observation times $t_0, t_1, \ldots, t_k$ completely determine the model.

More explicitly, if we can compute the covariances $C^{\alpha}_{ij} = \int_{t_{k-1}}^{t_k} \Sigma_{uT_i}\Sigma_{uT_j} du$, then the discount bond prices can be represented as

$$\frac{P_{t_\alpha T_i}}{B_t} = P_{0T_i} \exp\left(\sum_{\beta=1}^{\alpha} X^i_\beta - \frac{1}{2}\sum_{l=1}^{k} C^\beta_{ii}\right)$$

where the X^i_α are Gaussian random variables such that $\mathrm{Cov}(X^i_\alpha, X^j_\alpha) = C^\alpha_{ij}$ and $\mathrm{Cov}(X^i_\alpha, X^j_\beta) = 0$ for $\alpha \neq \beta$. As all derivative valuations can be reduced to expectations of functions of the discount bond prices, this gives a direct recipe for a Monte Carlo simulation of the model. In general, however, a representation of the model on a numerical grid is not efficiently possible, as the number of discount bonds one needs to represent is typically fairly large, and Monte Carlo methods are typically more efficient than grid methods in dimension four and above.

5.5.2 Markov Methods

The reason that a general Gaussian HJM model needs a large number of factors is that under a general HJM volatility term structure the discount bond prices become decorrelated over time. If the term structure is separable, however (compare appendix A), then this is not the case, and a far smaller number of factors is needed. For the generalised Vasicek model we saw earlier that a single-factor is sufficient.

This not only gives a far lower-dimensional Monte Carlo simulation, but it also shows that a low-dimensional grid (or tree) can be used to compute prices as expectations. The Markov functional grid methods we discussed for the one-dimensional case generalise directly to higher-dimensional ones. For example, a two-dimensional grid (or tree) suffices to value contingent claims in the two-factor generalised Vasicek model. Typically, Monte Carlo methods are more efficient in dimension four or higher.

6 LIBOR and Swap Market Models

Most interest rate derivatives depend on a finite number of points on the yield curve at a discrete set of times. To price them, we only need to understand the dynamics of a corresponding finite number of rates between these discrete times.

Market Models use a set of forward LIBORs (LIBOR Market Model or LMM) or a set of coterminal swap rates (Swap Market Model or SMM) as the basic quantities to propagate. The LIBOR Market Model was first introduced by Brace, Gatarek and Musiela in *Brace et al.* [1997]. The Swap Market Model was first described by Jamshidian in *Jamshidian* [1997]. Both used log-normal dynamics for underlying forward rates. They are consistent with the HJM framework.

The advantages of market models stem from the fact that the quantities modelled are directly observable market rates. Hence they can be modelled in line with market convention, e.g. as log-normal, and their calibration to traded instruments is automatic. Also, the sensitivity of the model to market prices is transparent, and alternative dynamics for the market rates have a straightforward representation in the model.

On the other hand, market models have the disadvantage that the model gives no information about the yield curve at times other than the selected ones. Also, the dynamic assumptions for different types of market models are typically not quite consistent. For example, forward LIBORs and forward swap rates cannot be simultaneously log-normal in an arbitrage-free market. Hence a log-normal LIBOR Market Model and a log-normal Swap Market Model cannot be consistent. Nor can LIBOR market models with different LIBOR frequencies. The differences are small, however.

A practical disadvantage is that market models are inherently high-dimensional.

6.1 The LIBOR Market Model

We assume $T_0 < T_1 < \cdots < T_n$ are the times relevant to chosen derivative, that F_i is the forward rate accruing from T_{i-1} to T_i, that its reset date is $t_i \leq T_{i-1}$ and that α_i is the corresponding day count fraction. We also denote the price at time t of the discount bond expiring at time T by P_{tT}.

Example. *Consider a ten-year semi-annual Bermudan swaption, exercisable on each reset date, starting on 1 June 2007. The times T_i are the value dates of the LIBORs referenced plus the final payment date, i.e. the first working day on or after the first of June and the first of December for each year from 2007 to 2016 (inclusive).*

N.B. *For simplicity of notation, textbooks often do not take account of spot days, i.e. for each i T_{i-1} is treated as both the reset date and the beginning of the accrual period of F_i and T_i is assumed to be the end of the accrual period as well as the maturity date of F_i. In an actual model implementation one should carefully distinguish between these. However, one needs to use the LIBOR with spot date T_{i-1} to calibrate the dynamics of F_i.*

The forward rate F_i must be a martingale in the $P_{\cdot,T_i}$-measure, and the price at time t of a FRA from T_{i-1} to T_i is $\alpha_i F_i(t) P_{tT_i}$, independent of the dynamics we assume for the forward rates.

The simplest natural model is to assume log-normal dynamics for each forward rate, as that is the implicit assumption underlying volatility quotes in the caplet market, i.e.

$$\frac{dF_i(t)}{F_i(t)} = \sigma_i^{\ln}(t)dW_t^i + \mu_i^{\ln}(t)dt$$

We will use a slight generalisation here and assume that the forward rates are *displaced* log-normal with displacement δ_i, so that

$$\frac{dF_i(t)}{F_i(t) + \alpha} = \sigma_i(t)dW_t^i + \mu_i(t)dt \tag{41}$$

The displacement is a simple way to model smiles and imposes little extra effort.

In this model the Black formulae hold for the displaced forwards $F_i + \delta_i$, hence the price of a caplet on F_i struck at K is

$$\mathrm{Caplet_i(t)} = P_{tt_i}\,\mathrm{Black}(F_i(t) + \delta_i, K + \delta_i, \bar{\sigma}_i, t_{i-1})$$

where

$$\mathrm{Black}(F, K, \sigma, t) = F\mathcal{N}\left(\frac{\log\frac{F}{K} + \frac{1}{2}\sigma^2 t}{\sigma\sqrt{t}}\right) - K\mathcal{N}\left(\frac{\log\frac{F}{K} - \frac{1}{2}\sigma^2 t}{\sigma\sqrt{t}}\right)$$

is the Black formula for caplets, $\mathcal{N}$ is the cumulative normal distribution and $\bar{\sigma}_i$ is the *term* volatility of $F_i + \delta_i$.

Hence we can easily calibrate the $\bar{\sigma}_i$ from quoted implied caplet volatilities $\hat{\sigma}_i$. As $\mathcal{N}$ is approximately linear near 0, this simplifies for at-the-money caplets to the approximation

$$\bar{\sigma}_i = \frac{F_i(t)}{F_i(t) + \delta_i}\hat{\sigma}_i$$

Having specified the stochastic drivers of the forward rates, the drifts are determined by no arbitrage. To compute them, we need to choose a pricing measure to work in. The natural choice is to use one of the discount bonds $P_{\cdot,T_i}$ ($i = 1, \ldots, n$), $P_{\cdot,T_k}$, say. We now compute the drift μ_i^k of F_i in this measure for all i.

As the FRAs a tradable, $\alpha_i F_i \frac{P_{\cdot,T_i}}{P_{\cdot,T_k}}$ must be a martingale for each i. In particular, F_k is a martingale, i.e. $\mu_k^k \equiv 0$. Furthermore, as zero coupon bonds are tradable, $\frac{P_{\cdot,T_i}}{P_{\cdot,T_k}}$ is also a martingale.

Hence, comparing drifts in

$$d\left(F_i\frac{P_{\cdot,T_i}}{P_{\cdot,T_k}}\right) = \frac{P_{\cdot,T_i}}{P_{\cdot,T_k}}dF_i + F_i d\left(\frac{P_{\cdot,T_i}}{P_{\cdot,T_k}}\right) + d\left\langle F_i, \frac{P_{\cdot,T_i}}{P_{\cdot,T_k}}\right\rangle$$

we obtain

$$\mu_i^k(t) = -\frac{1}{F_i(t)+\delta_i}\frac{P_{tT_k}}{P_{tT_i}} d\left\langle F_i(t), \frac{P_{tT_i}}{P_{tT_k}}\right\rangle$$

For $i < k$

$$\frac{P_{tT_i}}{P_{tT_k}} = \prod_{j=i+1}^{k}\left(1+\alpha_j F_j(t)\right)$$

so

$$d\left(\frac{P_{tT_i}}{P_{tT_k}}\right) = \frac{P_{tT_i}}{P_{tT_k}}\sum_{j=i+1}^{k}\frac{\alpha_j}{1+\alpha_j F_j(t)}dF_j(t) + (\cdots)dt.$$

Hence

$$\begin{aligned}\mu_i^k(t) &= -\frac{1}{F_i(t)+\delta_i}\sum_{j=i+1}^{k}\frac{\alpha_j}{1+\alpha_j F_j(t)}\left\langle F_i(t), F_j(t)\right\rangle \\ &= -\sum_{j=i+1}^{k}\frac{\alpha_j(F_j(t)+\delta_i)}{1+\alpha_j F_j(t)}\rho_{ij}\sigma_i(t)\sigma_j(t)\end{aligned}$$

where ρ_{ij} is the correlation between W^i and W^j.

For $i > k$

$$\frac{P_{tT_i}}{P_{tT_k}} = \prod_{j=k+1}^{i}\frac{1}{1+\alpha_j F_j(t)}$$

so

$$d\left(\frac{P_{tT_i}}{P_{tT_k}}\right) = -\frac{P_{tT_i}}{P_{tT_k}}\sum_{j=k+1}^{i}\frac{\alpha_j}{1+\alpha_j F_j(t)}dF_j(t) + (\cdots)dt.$$

Hence

$$\begin{aligned}\mu_i^k(t) &= \frac{1}{F_i(t)+\delta_i}\sum_{j=k+1}^{i}\frac{\alpha_j}{1+\alpha_j F_j(t)}\left\langle F_i(t), F_j(t)\right\rangle \\ &= \sum_{j=k+1}^{i}\frac{\alpha_j(F_j(t)+\delta_i)}{1+\alpha_j F_j(t)}\rho_{ij}\sigma_i(t)\sigma_j(t)\end{aligned}$$

where ρ_{ij} is again the correlation between W^i and W^j.

Summing up,

$$\mu_i^k(t) = \epsilon_{ik}\sum_{\min(i,k)<j\le\max(i,k)}\frac{\alpha_j(F_j(t)+\delta_i)}{1+\alpha_j F_j(t)}\rho_{ij}\sigma_i(t)\sigma_j(t) \tag{42}$$

where

$$\epsilon_{ik} = \begin{cases} -1 & \text{if } i < k \\ 0 & \text{if } i = k \\ 1 & \text{if } i > k \end{cases}$$

Note that the drift is *state-dependent*, i.e. it depends on the *stochastic* processes $f_i(t)$.

6.2 The Swap Market Model

Keeping the same notation as before, let us now consider instead of LIBOR forward rates the co-terminal swap rates. Let $S_i(t)$ be the market rate at time t of a swap that starts at time T_{i-1} and pays at times $T_i, \ldots, T_n$. S_i resets at t_i, and the set $\{S_1(t), \ldots, S_n(t)\}$ contains the same information about the yield curve as the set of LIBOR forward rates $\{F_1(t), \ldots, F_n(t)\}$.

The fixed leg of the swap pays $S_i(t)B_i(t)$, where

$$B_i(t) = \sum_{j=i}^{n} \alpha_i P_{tt_i}.$$

$B_i(t)$ is the *swap annuity* or the *basis point value* of a swap.

Again, we will assume that the rates we model follow a displaced log-normal process

$$\frac{dS_i(t)}{S_i(t) + \delta_i} = \sigma_i(t)dW_t^i + \mu_i(t)dt. \tag{43}$$

As before, no-arbitrage conditions determine the drifts. Let us first consider the numeraire P_{tT_n}, which is equivalent to using $B_n(t)$, and compute the drifts μ_i^n of the S_i in the corresponding measure. As both the swap and the swap annuity are tradable, $S_i\frac{B_i}{P_n}$ and $\frac{B_i}{P_n}$ are both martingales in this measure, hence comparing drifts in

$$d\left(S_i\frac{B_i}{P_{\cdot,T_n}}\right) = \frac{B_i}{P_{\cdot,T_n}}dS_i + S_i d\left(\frac{B_i}{P_{\cdot,T_n}}\right) + d\left\langle S_i, \frac{B_i}{P_{\cdot,T_n}}\right\rangle$$

we obtain

$$\mu_i^n(t) = -\frac{1}{S_i(t) + \delta_i}\frac{P_{tT_n}}{B_i(t)}\left\langle S_i(t), \frac{B_i(t)}{P_{tT_n}}\right\rangle.$$

As $S_{i+1}B_{i+1} = P_{\cdot,T_i} - P_{\cdot,T_n}$ for $0 \leq i < n$ we have

$$\begin{aligned}\frac{B_i(t)}{P_{tT_n}} &= \frac{\alpha_i P_{tT_i} + B_{i+1}(t)}{P_{tT_n}} = \alpha_i\left(S_{i+1}(t)\frac{B_{i+1}(t)}{P_{tT_n}} + 1\right) + \frac{B_{i+1}(t)}{P_{tT_n}} \\ &= \alpha_i + (1 + \alpha_i S_{i+1}(t))\frac{B_{i+1}(t)}{P_{tT_n}}\end{aligned}$$

hence, by recursion,

$$d\left(\frac{B_i(t)}{P_{tT_n}}\right) = \alpha_i \frac{B_{i+1}(t)}{P_{tT_n}} dS_{i+1}(t) + (1 + \alpha_i S_{i+1}(t)) d\left(\frac{B_{i+1}(t)}{P_{tT_n}}\right) + (\cdots)dt$$
$$= \frac{B_i(t)}{P_{tT_n}} \sum_{j=i+1}^{n} \eta_{ij}(t) dS_j(t) + (\cdots)dt.$$

where $\eta_{i,i+1}(t) = \alpha_i \frac{B_{i+1}(t)}{B_i(t)}$ and $\eta_{ij}(t) = (1 + \alpha_i S_{i+1}(t)) \frac{B_{i+1}(t)}{B_i(t)} \eta_{i+1,j}(t)$ for $j > i + 1$.

It follows by induction that for $i < j$

$$\eta_{ij}(t) = \alpha_{j-1} \frac{B_j(t)}{B_i(t)} \left(\prod_{i \le k < j-1} (1 + \alpha_k S_{k+1}(t)) \right).$$

We obtain

$$\mu_i^n(t) = -\sum_{j=i+1}^{n} \eta_{ij}(t)(S_j(t) + \delta_i)\sigma_i(t)\sigma_j(t)\rho_{ij},$$

where ρ_{ij} again denotes the correlation between W^i and W^j.

Under the numeraire B_k, we obtain for the drift of S_i by Girsanov's theorem

$$\mu_i^{B_k}(t) = \mu_i^n(t) + \frac{1}{S_i(t) + \delta_i} \frac{P_{tT_n}}{B_k(t)} \left\langle S_i(t), \frac{B_k(t)}{P_{tT_n}} \right\rangle$$
$$= \mu_i^n(t) + \sum_{j=k+1}^{n} \eta_{kj}(t)(S_j(t) + \delta_i)\sigma_i(t)\sigma_j(t)\rho_{ij}.$$

For the numeraire $P_{\cdot,T_k} = \frac{1}{\alpha_k}(B_k - B_{k+1})$, $k < n$, calculate

$$d\left[\frac{P_{\cdot,T_k}}{P_{\cdot,T_n}}\right] = \frac{1}{\alpha_k}\left(\frac{B_k}{P_{\cdot,T_n}} \sum_{j=k+1}^{n} \eta_{kj} dS_j - \frac{B_{k+1}}{P_{\cdot,t_n}} \sum_{j=k+2}^{n} \eta_{k+1,j} dS_j \right) + (\cdots)dt$$
$$= \frac{B_{k+1}}{P_{\cdot,T_n}} \left(dS_{k+1} + S_{k+1} \sum_{j=k+2}^{n} \eta_{k+1,j} dS_j \right) + (\cdots)dt$$

The drift μ_i^k of S_i in the corresponding measure is hence, by Girsanov's theorem,

$$\mu_i^k(t) = \mu_i^n(t) + \frac{1}{S_i(t) + \delta_i} \frac{P_{tt_n}}{P_{tt_k}} \left\langle S_i(t), \frac{P_{tt_k}}{P_{tt_n}} \right\rangle$$
$$= \mu_i^n(t) + \frac{B_{k+1}(t)}{P_{tt_k}} (S_{k+1}(t) + \delta_i)\sigma_i(t)\sigma_{k+1}(t)\rho_{i,k+1}$$
$$+ \frac{B_{k+1}(t)}{P_{tt_k}} S_{k+1}(t) \sum_{j=k+2}^{n} (S_j(t) + \delta_i)\eta_{k+1,j}(t)\sigma_i(t)\sigma_j(t)\rho_{ij}.$$

6.3 Consistency With the HJM Framework

The main contribution of Brace, Gatarek and Musiela in *Brace et al.* [1997], which Jamshidian quickly generalised to swap-based market models in *Jamshidian* [1997], was that market models are consistent with the HJM framework, thus showing that they satisfy all the main criteria for an interest rate model.

The idea behind this is quite straightforward. First, one extends the LIBOR market model to a model of the whole forward curve by prescribing dynamics for the discount bonds in between the discount bonds defined by the LIBOR term structure. A natural choice is linear interpolation, i.e. for $T_i \leq T \leq T_{i+1}$ and numeraire bond $P_{\cdot T_k}$ one defines

$$\frac{P_{tT}}{P_{tT_k}} = \frac{T_{i+1} - T}{T_{i+1} - T_i} \frac{P_{tT_i}}{P_{tT_k}} + \frac{T - T_i}{T_{i+1} - T_i} \frac{P_{tT_{i+1}}}{P_{tT_k}}$$

which makes the numeraire-adjusted discount bond expiring at T a martingale in the T_k-forward measure. One then shows that the HJM conditions are satisfied for this model.

7 Implementation Methodologies for Market Models

Leaving a detailed discussion of calibration for later, assume that we have all the data necessary to describe a LIBOR market model. How do we compute the price of a derivative?

- For a choice of numeraire N, the price of a derivative D at time t fulfils $D(t) = N_t \mathbb{E}_N[\frac{D}{N}|\mathcal{F}_t]$, where $\mathbb{E}_N$ denotes expectation in the pricing measure associated to N.
- By our initial assumptions for market models, the measure is determined by the distribution of the forward rates F_i at the reset times $t_0, \ldots, t_{n-1}$. As F_i resets at t_{i-1}, this is a $\frac{n(n+1)}{2}$-dimensional distribution
- Because of the high dimensionality and the state-dependence of the drifts, a Monte Carlo simulation is the most viable choice of implementation.

The most straightforward Monte Carlo implementation is to use an Euler scheme. This works as follows for a step from t to $t + \delta$:

1. Given $F_i(t)$ for all i, calculate the instantaneous drifts $\mu_i(t)$ at t and estimate the total drift by $\bar{\mu}_i = \mu_i(t)\delta$.

2. Draw a set of n independent normal random variables Z_i.

3. Compute the term covariance matrix

$$C = (c_{ij}) = \left(\int_t^{t+\delta} \sigma_i(t)\sigma_i(t)\rho_{ij} dt \right)$$

and a pseudo square root A of C, i.e. $AA^t = C$.

4. Estimate the values of the forward rates at time $t + \delta$ by

$$\log f_i(t + \delta) = \log f_i(t) + \bar{\mu}_i - \frac{c_{ii}}{2} + \sum_{j=1}^{n} a_{ij} Z_j$$

If the drifts $\mu_i(t)$ were constant, this would be exact, in the sense that the forward rates at time $t + \delta$ would be sampled with the correct distribution. If the drifts were deterministic, using $\bar{\mu}_i = \int_t^{t+\delta} \mu_i(t)dt$ in the above would also give an exact estimate. However, the state-dependence of the drifts introduces a *discretisation error* into the computation, and the step size has to be made fairly small to ensure a good approximation. The Euler scheme hence has a high computational cost.

A better estimation of the drift is given by the *predictor-corrector method*. This works as follows:

1. Estimate the forward rates at time $t + \delta$ by the Euler scheme.

2. Compute the drift $\mu_i(t + \delta)$ with the estimated forward rates.

3. Repeat steps 2–4 using the average of the two instantaneous drifts $\mu_i(t)$ and $\mu_i(t + \delta)$ in place of $\bar{\mu}_i$.

The predictor-corrector method has proven to be very accurate for time steps as long as five years, see e.g. *Hunter et al.* [2001]. Consequently, we can choose the simulation time steps to coincide with the reset times of the underlying rates in most cases.

The predictor-corrector method is an example of a higher-order estimation scheme; for a detailed discussion of such schemes see *Kloeden and Platen* [1992]. There have been investigations of other schemes, but the predictor-corrector seems to be the preferred choice in practice.

There are other aspects of the algorithm which are interesting for an efficient implementation. We have not discussed the pseudo-square root appearing in the Euler scheme yet. The covariance matrix C for a time interval is non-negative symmetric, and such a matrix has many pseudo-square roots — right multiplication of a pseudo-square root by a unitary gives another pseudo-square root, and all pseudo-square roots arise this way.

The simplest way to compute a pseudo-square root is *Cholesky decomposition*, which gives a lower-diagonal matrix. When we use a pseudo-random number generator, this is a satisfactory choice. It is also numerically the most efficient, since multiplication with a triangular matrix requires fewer operations than multiplication with a general pseudo-square root.

N.B. *As the same covariance matrix is used for each simulation of a given time step, the pseudo-square root need only be computed once, in the set-up phase of the simulation.*

Pseudo-random numbers give convergence of order $N^{-\frac{1}{2}}$, where N is the number of samples drawn. The convergence order can be improved by using quasi-random number sequences, such as Sobol, Fauré or Niederreiter sequences, to almost order N^{-1}.

However, it is not enough simply to replace a pseudo-random number sequence by a quasi-random number sequence. Quasi-random number sequences are constructed to systematically sample the range of the random variables they represent. This sampling is much more effective for low-dimensional problems than for high-dimensional ones, and high-dimensional sequences sample the dimensions unevenly — the first coordinates are much better sampled than the last ones.

Hence the choice of pseudo-square roots can influence the *convergence coefficient* considerably.

A covariance matrix C has eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Let Λ be the diagonal matrix with entries $\sqrt{\lambda_i}$. Then there exists an orthogonal matrix U such that $A = U\Lambda$ is a pseudo-square root of C. This is the most effective pseudo-square root for use with a quasi-random number sequence.

The computation of the eigenvalue decomposition of a covariance matrix has the same order of complexity as a Cholesky decomposition and increases with the third power of the dimension. However, it is typically still considerably more expensive than a Cholesky decomposition to compute.

Now consider the construction of a Brownian motion W at time $0 = t_0 < t_1 < \cdots < t_n$ from independent normal random variables Z_i. As the increments of a Brownian motion are independent, we could use $W(t_i) - W(t_{i-1}) = \sqrt{t_i - t_{i-1}} Z_i$.

A quasi-random number sequence will sample the Brownian motion much better at the initial time steps than at the later ones. If the final state of the Brownian motion is of more interest to us than the intermediate states, we should use the first dimension to sample the total time step. A *Brownian bridge* gives the conditional distribution of the intermediate state of a Brownian motion. More precisely, for $0 < t_1 < t_2$, the distribution of W_{t_1} conditional on $W_{t_2} = y$ is $\frac{t_1}{t_2}y + \sqrt{t_1(1 - t_1/t_2)}N(0, 1)$.

The dimensionality of the random vector we need to draw for each simulation path is equal to the sum of the number of live forward rates for each time step.

The best way to use quasi-random numbers is to combine Brownian bridging between the time steps with spectral decomposition on the covariance matrix in each single time step.

The optimal way to combine the two may depend on the derivative in question and the volatility term structure.

A good standard Brownian bridge is to use the first dimension for the final time, the second for the mid-point, and then to use the further dimensions to repeatedly sub-divide the time intervals.

With pseudo-random numbers these techniques at worst do not affect the performance of the Monte Carlo simulation, so they may represent an overhead. However, Brownian bridging at least is typically beneficial.

7.1 Efficient Design of Monte Carlo Frameworks

We have been discussing how to create efficient Monte Carlo simulations of the basic quantities that describe an interest rate market, i.e. the numeraire adjusted discount factors. The efficient and flexible application to derivative payoffs requires careful additional design.

The design most Monte Carlo implementations initially seem to adopt is the following:

- Set up Monte Carlo simulation of the stochastic process, including time steps, to achieve desired mathematical properties.
- Create a "market" for each time step that permits products to get discount factors and other simulated market data.
- Define the mapping between the "markets" created and the times/dates they correspond to.

Since explicitly expressing every payoff in terms of discount factors is complicated and error prone, not least because of the various market conventions, the "market" will typically provide functions to compute reference rates such as LIBOR or vanilla swap rates.

Payoffs are then priced by calculating cash payment amounts from the markets corresponding to fixing dates of rates determining the behaviour of the payoff. The cash payment amounts are then converted to numeraire units and discounted.

While this approach is natural from a mathematical perspective, it creates substantial inefficiencies in a pricing system:

- Computation of market reference rates from the “market” requires interpretation of conventions on demand, in every Monte Carlo path, or buffering within the “market”

- Handling the mapping of dates to “market” states in the payoff complicates the payoffs and is error prone

A better approach starts from the realisation that pricing Monte Carlo simulations for specific payoffs is defined by the *observations* the payoffs need to make during the simulation to determine the cash flows that they pay. An *observation* is defined by the time at which it is to be made and by a description of the *observable*, i.e. the market quote.

Both the observations and the rules to determine the cash flows are known up front. That is the basic information provided by term sheets for the payofs. Furthermore, we can distinguish *observables* by whether they are can be calculated from other, simpler *observables*. Such *composite observables* do not need to be implemented by a model but can be implemented generally for all models. Only *observables* that cannot be decomposed, which we call *atomic observables*, need to be provided by a model.

In the interest rate context, the numeraire adjusted discount bond prices are the natural *atomic observables*. All other interest rates can be calculated from them, since they describe the full interest rate curve and provide all the information necessary to discount cash flows.

This suggests the following design approach:

- Provide a standard way to describe all observables that a payoff can use. This can be a hierarchy of classes providing a standard interface in an object oriented language such as C++, or a domain specific description language.

- Require all payoffs to specify the observations they want to make and the cash flows they can generate (characterised by determination date, payment date and currency, *not* the amount) within the simulation *before* the simulation is set up. A unique identifier is allocated to every observation.

- Each observation is allocated a unique handle, and discount factor observations to discount cash flows are generated automatically as required.

- The underlying observations for a composite observation are created automatically and added to the set of observations required. This is carried out recursively until all atomic observations required to compute all composite observations are known.

- The atomic observations are passed to a model before the calibration is called and the simulation created. The model can use the information to configure the simulation optimally. The model must provide values for all atomic observations for all scenarios generated.

- Given a set of atomic observations, all composite observations are calculated.

- The set of values for all observations is passed to the payoffs, which can use the identifiers provided during the specification of the observations to obtain the numerical value for each observation they reference. The payoffs use this to compute their cash flows.

- The cash flows are post-processed, e.g. to calculate PV.

Note that information about market data and market conventions is only required during the initial specification phase. The calculations required for each Monte Carlo path can be fully configured and stripped of any look-up or computation related to market conventions. The values for the observations can be passed in arrays or similarly efficient memory structures, and the identifiers can be simple indices for look-up within the array. The resulting code is typically very fast and efficient.

The design sketched here is an elaborate example of three general principles that are useful in designing efficient numerical calculations:

1. Compile rather than interpret, i.e. take as much information as possible into account when setting up a computation and take decisions as early as possible.

2. Abstract and simplify, i.e. minimise the information required to be passed between different layers of a calculation.

3. Systematically avoid unnecessary computations, such as repeated computations of the same value.

7.2 Risk Sensitivity Calculations

We now briefly discuss the calculation of risk sensitivities within a Monte Carlo simulation. The simplest way to perform risk sensitivity calculations is to use finite difference approximation. We now analyse the properties of finite difference approximations on Monte Carlo simulations.

7.2.1 Finite Difference

For background recall the properties of the basic finite difference approximation formulae:

- The forward approximation of the first order derivative of a function $f(x)$ is

$$f'(x) \approx \Delta_{x,h} f(x) = \frac{f(x+h) - f(x)}{h} \tag{44}$$

  for which the optimal size of h is proportional to $\sqrt{\left|\frac{\epsilon_f f(x)}{f''(x)}\right|}$ where ϵ_f is the fractional accuracy of f. The fractional accuracy of $\Delta_{x,h} f(x)$ is of the order $\sqrt{\epsilon_f}$ with this choice of h.

- The central finite difference approximation is

$$f'(x) \approx \bar{\Delta}_{x,h} f(x) = \frac{f(x+h) - f(x-h)}{2h} \tag{45}$$

  for which the optimal size of h is proportional to $\left|\epsilon_f \frac{f(x)}{f'''(x)}\right|^{1/3}$ and the resulting fractional accuracy is of the order $\epsilon_f^{2/3}$.

- The central finite difference approximation of the second order derivative of f is given by

$$f''(x) \approx \Delta^2_{x,h} f(x) = \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}$$

  The optimal scaling of h is $\sim \epsilon_f^{1/4} \left|\frac{f(x)}{f'''(x)}\right|^{1/3}$ and fractional accuracy is of the order $\sqrt{\epsilon_f}$.

The generic accuracy of a Monte Carlo estimation is proportional to $N^{-1/2}$ where N is the number of Monte Carlo samples (or paths). Hence the optimal bump size h for the central finite difference estimator is of order $N^{-1/6}$ and the standard error of the resulting approximation is of the order $N^{-1/3}$. The forward finite difference approximation converges even more slowly, with standard error of the order $N^{-1/2}$ for optimal bump size of the order $N^{-1/4}$. The second order finite difference approximation converges also with standard error of the order $N^{-1/4}$ for optimal bump size of the order $N^{-1/8}$.

It is possible to improve upon this by using common random samples, i.e. by using the same random samples in each function evaluation within the finite difference calculation. The Monte Carlo estimator is the expectation of a random variable $V(\xi)$ dependent on the parameter ξ that we want to differentiate with respect to. If expectation and differentiation can be interchanged and V is differentiable (almost surely) with respect to ξ, then

$$\frac{\partial \mathbb{E}[V(\xi)]}{\partial \xi} = \mathbb{E}\left[\frac{\partial V(\xi)}{\partial \xi}\right] \tag{46}$$

and the derivative can be computed as a Monte Carlo simulation which converges with $N^{-1/2}$. The finite difference calculation can then be done "path-by-path" as

$$\Delta_{\xi,h}\mathbb{E}[V(\xi)] = \mathbb{E}\left[\Delta_{\xi,h}V(\xi)\right] \to \mathbb{E}\left[\frac{\partial V(\xi)}{\partial \xi}\right] \tag{47}$$

for $h \to 0$. One consequence is that

$$\operatorname{Var}\left(V(\xi + h) - V(\xi)\right) \sim h^2 \operatorname{Var}\left(\frac{\partial V(\xi)}{\partial \xi}\right) \tag{48}$$

and there is no variance penalty for choosing h to be small as there is for the generic case above.

The bump size h can be chosen on the basis of the accuracy of the calculation of V on a single path. This will typically be machine accuracy unless the dependency of V on the parameter ξ is via the calibration of the model, which will typically have lower accuracy. The bias will hence be vanishingly small relative to the Monte Carlo error, and the estimator will converge with the canonical Monte Carlo convergence rate of $N^{-1/2}$.

Expectation and differentiation can be interchanged if the difference quotients of V with respect to ξ are uniformly integrable; the standard condition for this to be valid is that V is almost surely differentiable with respect to ξ and satisfies a Lipschitz condition almost surely, see e.g. [*Asmussen and Glynn*, 2007, Ch. VII, Proposition 2.3]. One can also check equation (48) numerically to test whether this condition is met.

As a general rule of thumb, the condition of interchangeability of expectation and differentiation is met if the payoff function is a continuous and almost everywhere continuously differentiable function of the market observables it references and these observables are smooth functions of the parameter ξ with respect to which the sensitivity is computed. For the standard first order risk sensitivities (Deltas/PV01s and Vegas) this will typically be the case if the payoff does not have discontinuities. However, many payoffs in Finance are discontinuous, and hence the method above is not universally applicable.

If equation (46) does not hold, then the convergence behaviour of finite difference schemes with common random samples is worse than $N^{-1/2}$. The use of common random samples is still beneficial since unlike the generic case one expects $\operatorname{Var}\left(V(\xi + h) - V(\xi)\right) \to 0$ for $h \to 0$; however, the order of convergence is typically only h instead of h^2 when equation (46) holds.

In this case the standard error of the first order finite difference quotients is $\sim (hN)^{-1/2}$, hence the optimal order of convergence is $N^{-1/3}$ for $h \sim N^{-1/3}$ for the forward finite difference estimator, and $N^{-2/5}$ for $h \sim N^{-1/5}$ for the central finite difference estimator. For the second order finite difference quotients the Monte Carlo error is $\sim h^{-3/2}N^{-1/2}$, hence the optimal order of convergence is $N^{-2/7}$ for $h \sim N^{-1/7}$.

These convergence rates are better than those for general Monte Carlo, so common random samples should generally be used. However, while the first order central finite difference estimator is not too far from the optimal rate of convergence for Monte Carlo, smoothing the payoff to take advantage of (46) will generally be advantageous, particularly since a determination of the optimal bump size $h \sim N^{-1/5}$ will not be required to obtain a good rate of convergence. For second order only a very small improvement over the generic convergence rate can be achieved with common random samples, so smoothing or other techniques to should generally be used.

7.2.2 Pathwise Differentiation and Adjoint Calculation

If (46) holds then the obvious alternative to finite difference approximation is explicit differentiation of the payoff function, i.e. Monte Carlo simulation of the derivative as in the right hand side of (46). This is straightforward in principle but does not necessarily simplify the computation significantly, since the forward finite difference approximation on a path-by-path basis is quite accurate, particularly when compared with the Monte Carlo error, as we saw above.

However, using an *adjoint* calculation of risk sensitivities can significantly improve efficiency. The principle behind adjoint sensitivity calculation is simple. The value of a payoff on a Monte Carlo path is a composition of various functions that map the original calibration parameters to Monte Carlo sample of a stochastic process to atomic observables (e.g. numeraire adjusted discount factors) to composite observables to cash flows to present value. Symbolically,

$$V(x_1, \ldots, x_n) = f_k \circ f_{k-1} \circ \cdots \circ f_1(x_1, \ldots, x_m)$$

where the x_i are the parameters to which we want to compute sensitivities. By the chain rule the total derivative is

$$DV(x_1, \ldots, x_n) = Df_k(f_{k-1} \circ \cdots \circ f_1(x_1, \ldots, x_m))Df_{k-1}(\cdots) \cdots Df_1(x_1, \ldots, x_m))$$

where each of the individual (total) derivatives Df_i is a matrix. Matrix multiplication is associative, so once all the individual derivative matrices have been computed, then this can be calculated as

$$(\cdots((Df_k Df_{k-1})Df_{k-2})\cdots)\, Df_1(x_1, \ldots, x_m)$$

Since the value is a scalar, this is a sequence of vector-matrix multiplications rather than a sequence of matrix-matrix multiplications, which significantly reduces the computation time. Simultaneous computation of large numbers of sensitivities for a relatively small number of outputs (e.g. the value of a portfolio) can therefore be substantially accelerated using this adjoint method.

The cost of this is having to store the individual derivative matrices in memory before one can carry out the calculation. Depending on the number of parameters that can be costly compared to the normal memory use of the Monte Carlo path, but is unlikely to be problematic for a single Monte Carlo path on a modern computer.

Adjoint derivative calculation can easily be combined with the design of Monte Carlo frameworks discussed earlier. Composite observations, payoffs and discounting of cash flows are components in the chain of functions leading to the value of a derivative on a path, which can easily be differentiated. This leaves the derivatives of the atomic observables to be provided by the model.

8 Calibration of Market Models

We have so far glossed over the details of model calibration. As the discussion above has shown, we need to compute the covariance term structure of the underlying Brownian motions, or at least the term covariance matrices for the simulation time steps.

The data that we need for this are the volatility functions $\sigma_i(t)$ and the correlation matrix (ρ_{ij}). Let us address the volatility first.

The simplest thing to do would be to take the volatilities to be flat, i.e. $\sigma_i(t) \equiv \bar{\sigma}_i$, where $\bar{\sigma}_i$ is the market implied vol for the corresponding caplet. However, the caplet volatility market typically trades at a humped shape with the peak around the two-year point. If we keep volatilities flat, the market at future points in time will look very different.

To address this, we seek a *time-homogeneous* volatility term structure, i.e. a function σ so that $\sigma_i(t) = \sigma(t_{i-1} - t)$. This function has to match the market implied volatilities, i.e.

$$\bar{\sigma}_i^2 t_{i-1} = \int_0^{t_{i-1}} \sigma(t_{i-1} - t)^2 dt$$

There are many functions that can fulfil these equations. One way is to choose a parametrised family of functions and fit it.

An example that used to fit Western interest rate markets well and was popularised by Riccardo Rebonato is

$$\sigma(t; a, b, c, d) = (a + bt)\exp(-ct) + d$$

As the fit will in general not be perfect, we adjust by scaling factors K_i such that $\bar{\sigma}_i^2 t_{i-1} = K_i^2 \int_0^{t_{i-1}} \sigma(t_{i-1} - t; a, b, c, d)^2 dt$, and set $\sigma_i(t) = K_i \sigma(t_{i-1} - t; a, b, c, d)$.

The K_i should be close to 1 if our fit for $\sigma(\cdot; a, b, c, d)$ is good. If they are not, our market fit is bad.

Why is the choice of volatility term structure important?

- We may need to evolve a forward rate to a time other than its reset. To do that, we need to know its term volatility to that time.
- We may need to compute future values of caplets or other derivatives in our simulation. To do this, we need to know the future volatilities.
- The shape of the instantaneous volatility can cause decorrelation between forward rates — if two rates have different sensitivities to the same piece of information, this makes them less correlated.

We still need to specify the (instantaneous) correlation between the Brownian motions (and hence the rates). Correlation is not a directly observable quantity. Popular modelling choices are

- Exponential: $\rho_{ij} = e^{-\beta |t_{i-1} - t_{j-1}|}$
- Exponential with "long core": $\rho_{ij} = L + (1 - L) e^{-\beta |t_{i-1} - t_{j-1}|}$
- More general functions of the form $h(t_{i-1}, t_{j-1})$, e.g. $h(s, t) = \exp(-\beta \left| s^{0.5} - t^{0.5} \right|)$ or $h(s, t) = \left(\frac{\min(s,t)}{\max(s,t)} \right)^\beta$.
- Use swaption prices together with the caplet market to imply correlations. One must be careful, however, to use *stable* fitting algorithm.
- Use historical data to calibrate correlation matrices in regular intervals.

8.1 Calibration to Co-Terminal Swaptions

For some products, we will want to calibrate to the swaption market instead of the caplet market. Let

$$Z_{ij}(t) = \frac{F_i(t)}{S_j(t)} \frac{\partial S_j}{\partial F_i}(t),$$

then we have, up to drift terms,

$$d(\log S_i(t)) = \sum_{j=1}^{n} Z_{ij}(t) d(\log F_j(t)).$$

In fact, since

$$S_i(t) = \frac{P_{tT_0} - P_{tT_i}}{\sum_{j=1}^{i} \alpha_i P_{tT_j}}$$

and

$$\frac{P_{tT_i}}{P_{tT_0}} = \prod_{j=1}^{i} \frac{1}{1 + \alpha_i F_i}$$

we see that $Z_{ij}(t)$ is easily computed as a rational function of the F_i.

While forward LIBOR and forward swap rates cannot be simultaneously log-normal, it is nonetheless approximately true. Pretending that the S_i are log-normal with instantaneous covariance matrix $C_S(t)$ and denoting the covariance matrix of the F_i by $C_F(t)$, we have $C_S(t) = Z(t)C_F(t)Z(t)^t$.

We can use this for calibration by freezing coefficients, i.e. by using $Z \equiv Z(0)$ to relate LIBOR forward covariance matrices C_f and swap forward covariance matrices C_S by $C_S = ZC_F Z^t$. We calibrate as follows:

1. For each time t_{i-1} compute the forward LIBOR term covariance matrix $C_F(0, t_{i-1})$.

2. Compute the implied swaption covariance $v_i = (ZC_F(0, t_{i-1})Z^t)_{ii}$.

3. Use the market swaption vol $\bar{\sigma}_i$ to compute $\lambda_i = \frac{\bar{\sigma}_i}{\sqrt{\frac{v_i}{t_{i-1}}}}$.

4. Set Λ to be the diagonal matrix with entries λ_i.

5. For each LIBOR forward covariance matrix C_F set the corresponding swaption-calibrated covariance matrix to be

$$C'_F = Z^{-1}\Lambda(ZC_F Z^t)\Lambda(Z^t)^{-1} = (Z^{-1}\Lambda Z)C_F(Z^{-1}\Lambda Z)^t$$

This calibration to swaptions has the following properties:

- Calibration to swaption volatilities is automatic and stable.
- Prices are stable for changes in swaption volatilities.
- Swaption vegas can be computed easily.
- Inverting the transformation, we can also use this method to calibrate a swap market model.
- Calibrated this way, LIBOR and swap market models agree closely.

8.2 The Cascade Algorithm

While parametric forms of the term structure of volatility like the above can capture specific features very well and can be tailored to specific situations, they can be problematic to use in production systems. The optimisation of the parameters automatically can be unstable and several local minima of the optimisation problem may exist.

One alternative approach is the so-called *cascade algorithm*. It assumes:

- the volatility term structure of each LIBOR as a piecewise constant function
- the correlations between LIBORs are fixed and pairwise positive

It further uses the same approximations as above to create swaption and forward covariances, namely approximate simultaneous log-normality and freezing of the coefficient. This gives a quadratic relationship between the LIBOR and swaption volatilities.

Because of the piecewise constant volatility term structure for each LIBOR, this gives a quadratic formula relating the individual pieces of each LIBOR volatility term structure to a given swaption volatility.

Because the number of pieces for each LIBOR needed increases with the time to expiry, the calculation can be structured so that each swaption implied volatility determines one piece of one LIBOR term structure.

This assumes that the whole swaption volatility matrix is used; if this is not desirable because of data quality or because only a subset of swaptions is to be used for hedging, then constraints such as time homogeneity can be imposed instead.

Note that the algorithm relies on having a correlation structure available. This can then be re-fitted to the volatility fit, and one can achieve a calibration of both volatility and correlation term structures by iterative application of the fitting algorithms.

9 Markov Functional Methods and Models

Assume we have an arbitrage free (interest rate) market, numeraire N and a tradable asset A. Assume further that there exists a Markov process X (under the measure $\mathbb{P}_N$) such that the numeraire-adjusted price of A at some time t can be represented as a function of X_t:

$$\frac{A_t}{N_t} = F_A(t, X_t)$$

Then by the martingale property for $s < t$

$$\frac{A_s}{N_s} = K(s,t)F_A(t, X_t) \tag{49}$$

where $K(t, s)$ is the transition operator of X. $\frac{A_s}{N_s}$ is $\sigma(X_s)$ measurable, hence again a functional of X_s, $F_A(s, X_s)$. (Asset price) processes that can be represented this way are called *Markov Functionals*.

An interest rate model is of Markov Functional type if all zero coupon bond prices in a market can be represented as functionals of some Markov process X_t for all times t where it makes sense,

$$\frac{P_{tT}}{N_t} = F_{P_{\cdot T}}(t, X_t),$$

typically for $t \leq \min(T, T^*)$. This allows interest rates and (numeraire-adjusted) interest rate derivative prices to be represented as Markov functionals as well.

9.1 Pricing in a Markov Functional Model

By equation (49) above one "only" needs to understand the transition operator $K(s, t)$ of the Markov process to price derivatives in such a model,

$$F_A(s, X_s) = K(s, t)F_A(t, X_t).$$

In particular, the current price of the asset is given by transition to time zero,

$$A_0 = F_A(0, X_0) = K(0, t)F_A(t, X_t).$$

If $K(s, t)$ has a density $k(s, t; x, y)$, then

$$F_V(s, X_s) = \int k(s, t; X_s, y)F_V(t, y)dy. \tag{50}$$

When k is known explicitly, this can give very efficient calculation algorithms.

9.2 Examples

Short-rate models and affine models are obviously Markov functional; for example, in the Vasicek-Hull-White model the numeraire-adjusted discount bonds prices at a given time are exponential functions of a time-changed Brownian motion in any forward measure.

LIBOR or swap rate market models are completely described by the driving LIBOR or swap rate processes, which are typically Markov, so they can be described as Markov functional models.

More generally, if one restricts oneself to only a finite number of discount bond prices, which is permissible since cash flows can only occur at finitely many times in a typical derivative, then many HJM models (in particular all Gaussian ones) can be shown to have a driving Markov process, which may be very high-dimensional.

So Markov functional descriptions exist for basically all interest rate models of practical interest.

However, for this to be useful, the driving Markov process must have a transition kernel that allows an efficient numerical implementation. As the calculation will be based on functionals, these must be represented. In general, the complexity of this representation will increase exponentially with the dimension of the process, so only a low-dimensional driving Markov processes will be useful. Typically dimension three is the limit beyond which Monte Carlo methods become more efficient. Consequently, in common parlance "Markov Functional" implies that the driving Markov process is of dimension three or less.

In practice, one of the major concerns in pricing interest rate derivatives is the efficiency of the implementation. As low dimensional Markov Functional representations can allow very efficient implementations, models that admit these are of particular interest.

9.3 Markov Functional Models

One approach, suggested by *Hunt et al.* [2000], is to explicitly construct interest rate models as Markov Functional type, starting from a convenient low-dimensional Markov process. These models are called *Markov Functional models*.

A Markov Functional Model consists of a construction of the functionals representing the discount bond prices for the chosen set of dates (model calibration) together with a scheme to apply the transportation operator of the Markov process to functionals.

In the original paper *Hunt et al.* [2000] introducing Markov functional models, criteria for an interest rate model to be good for practical were that it

1. be arbitrage free,
2. be well-calibrated, i.e. price as many liquid instruments as possible without overfitting,
3. be realistic and transparent in its properties, and
4. allow an efficient implementation.

Markov functional models are arbitrage free and allow efficient implementation by construction. The functional representation is very flexible an can be used to obtain very good fits to vanilla instruments.This typically comes at a cost to transparency, though, and the behaviour of the model can be hard to understand or even become unrealistic ("mindless fitting").

The efficient implementation also has a cost; one is limited to low-dimensional Markov processes (up to dimension three, or maybe four) and is hence limited in the complexity of dynamics that can be represented. For most single-currency interest rate derivatives this need not be problematic, but it requires careful analysis for each new product.

While these limitations argue against Markov functional models as the mainstay solution for interest rate derivatives pricing, they are useful as efficient low-dimensional approximations to more complex models, which may be used to reduce computational load for particular products, and in particular in calibration problems.

In comparison to market models, they will typically be far more efficient to run, but require more implementation and analysis time from the quant team for each new product.

10 Implementation Techniques for Markov Functional Models

We will not discuss the pros and cons of Markov Functional Models any further but concentrate instead on the properties of the implementations. These are interesting in themselves; as we saw all commonly used interest rate models are of Markov Functional type, even though the underlying Markov process may be very high-dimensional.

Even if the model itself does not admit an efficient implementation as a Markov Functional Model, some simpler sub-pricing problems may do so. This is particularly relevant for calibration problems, where more efficient algorithms than Monte Carlo are highly desirable.

Even if the pricing of a calibration instrument cannot be exactly represented as a low-dimensional Markov Functional, it may be possible to use such a representation as an approximation.

In current practice, the Markov process driving a Markov functional model is almost exclusively Brownian motion, for which the transition density is the heat kernel

$$k(s,t;x,y) = \phi_n(x-y, t-s). \tag{51}$$

Here

$$\phi_n(x,t) = \frac{1}{(2\pi t)^{n/2}} \exp\left(-\frac{\|x\|^2}{2t}\right) \tag{52}$$

is the fundamental solution of the well-studied heat equation $\partial_t \phi_n = \frac{1}{2}\Delta\phi_n$ on $\mathbb{R}^n$, $\Delta = \sum_{i=1}^n \partial_{x_i}^2$.

The general pricing procedure is as follows:

1. Represent numeraire-adjusted value of all potential residual payoffs of the derivative at the final time t_n as Markov functional $V_{n,i}(X_{t_n})$ and set $k = n$.

2. Compute the Markov functional representation of the potential residual values at the next earlier time t_{k-1} using the transition formula

$$\tilde{V}_{k-1,i}(X_{t_{k-1}}) = \int V_{k,i}(y)k(t_{k-1}, t_k; X_{t_{k-1}}, y)dy$$

3. Combine the $\tilde{V}_{k-1,i}$ according to the term sheet to derive Markov functional representations $V_{k-1,i}$ of the residual values the derivative may take at t_{k-1}.

4. While $k > 1$ go to step 2 replacing k by $k - 1$.

At the final step we obtain the desired derivative value at time 0 as $V_0(X_0)$. Note that the first step hides the calibration procedure, which we will not explore further here.

10.1 Case Study: Markov Functional Approximation to an LMM

We now consider a specific example, a one-dimensional Markov Functional approximation to a standard LIBOR Market Model. We want to examine how well the approach performs and when it might break down. As we examined the numerical method earlier, in the context of the Vasicek-Hull-White model, we already know what the potential sources of numerical issues are:

- Approximation error in the piecewise polynomial representation, particularly in the wings.

- Both the interpolation scheme and the grid point placing are somewhat arbitrary.

- Smoothing only works in the direction of the Markov process, but not in the direction of state variables.

For the example model, we assume we have a term structure of LIBORs $L_0, \ldots, L_{n-1}$ spanning the $T_0 < \cdots < T_n$ and resetting at $t_0, \cdots, t_{n-1}$ (for further simplification may take $t_i = T_{i-1}$). We construct a Markov functional model in the terminal (T_n-forward) measure such that all LIBORs have log-normal law at expiry. That is, we require that

$$L_i(t_i) = \tilde{L}_i \exp\left(\gamma_i W_{c_i} - \frac{\gamma_i^2 c_i}{2}\right)$$

for some constants $\tilde{L}_i$, γ_i and c_i under the constraint $c_0 < \cdots < c_{n-1}$. We can regard c as a "clock function". For simplicity we assume that we can determine the γ_i and c_i directly from vanilla option prices.

Denote the value of the numeraire-adjusted discount bonds at the reset times by

$$\tilde{P}_{kl} = \frac{P_{t_k T_l}}{P_{t_k T_n}}$$

for $k \le l$. By definition, $\tilde{P}_{kn} \equiv 1$ for all k.

We need to construct Markov functional representations of $\tilde{P}_{kk}$ for $k \in \{0, \dots, n-1\}$. By definition of the LIBORs we have

$$\begin{aligned}\tilde{P}_{kk} &= (1 + \alpha_k L_k)\tilde{P}_{k,k+1} \\ &= \left(1 + \alpha_k \tilde{L}_k e^{\gamma_k W_{c_k} - \frac{\gamma_k^2 c_k}{2}}\right)\tilde{P}_{k,k+1}\end{aligned}$$

where α_k is the accrual fraction of L_k.

By induction $\tilde{P}_{k,k+1}$ is a functional of the Brownian motion W at clock time c_k, and calibration consists of determining $\tilde{L}_k$. Note that this can easily be extended to calibrate γ_k and c_k, but we'll avoid this for simplicity.

As $\mathbb{E}[\tilde{P}_{kk}] = \frac{P_{0T_k}}{P_{0T_n}} \equiv \tilde{P}_k$ is determined by the initial yield curve, it follows by taking expectations that

$$\tilde{L}_k = \frac{\tilde{P}_{0k} - \tilde{P}_{0,k+1}}{\alpha_k \mathbb{E}\left[\exp\left(\gamma_k W_{c_k} - \frac{\gamma_k^2 c_k}{2}\right)\tilde{P}_{k,k+1}\right]}$$

Clearly $\tilde{L}_{n-1}$ equals the forward LIBOR $L_{n-1}(0)$ as that LIBOR is a martingale in the terminal measure, but for the other LIBORs the simplification we made by approximating the drifts that would normally appear shows up in the $\tilde{L}_k$.

$$\begin{aligned}\tilde{L}_{n-2} &= \frac{\tilde{P}_{n-2} - \tilde{P}_{n-1}}{\mathbb{E}\left[e^{\gamma_{n-2} W_{c_{n-2}} - \frac{\gamma_{n-2}^2 c_{n-2}}{2}}\left(1 + \alpha_k \tilde{L}_{n-1} e^{\gamma_{n-1} W_{c_{n-2}} - \frac{\gamma_{n-1}^2 c_{n-2}}{2}}\right)\right]} \\ &= \frac{\tilde{P}_{n-2} - \tilde{P}_{n-1}}{1 + \alpha_k \tilde{L}_{n-1} \exp\left(\gamma_{n-2}\gamma_{n-1} c_{n-2}\right)} < L_{n-2}(0)\end{aligned}$$

In general, many more covariance terms appear:

$$
\begin{aligned}
\tilde{L}_k &= \frac{\tilde{P}_k - \tilde{P}_{k+1}}{e^{\gamma_k W_{c_k} - \frac{\gamma_k^2 c_k}{2}} \prod_{j=k+1}^{n-1} \left(1 + \alpha_j \tilde{L}_j e^{\gamma_j W_{c_j} - \frac{\gamma_j^2 c_j}{2}}\right)} \\
&= \frac{\tilde{P}_k - \tilde{P}_{k+1}}{\sum_{J \subseteq \{k+1,\ldots,n-1\}} \beta_J \exp(\gamma_k \gamma_J c_k)}
\end{aligned}
$$

where

$$
\beta_J = \mathbb{E}\left[\prod_{j\in J} \alpha_j \tilde{L}_j e^{\gamma_j W_{c_j} - \frac{\gamma_j^2 c_j}{2}}\right]
$$

and $\gamma_J = \sum_{j\in J} \gamma_j$. The β_J and γ_J can be computed by recursion on k, but this leads to a number of terms exponential in n, which quickly requires approximations.

However, if $\gamma_k \equiv \gamma$, then this simplifies drastically as $\gamma_J = \gamma_K$ for $|J| = |K|$. This allows a comparison to be made between exact analytical and numerical values for fairly realistic cases. Figure 1 shows some example values for $\alpha_k \equiv 0.5$, $\gamma_k \equiv 15\%$, $c_k = k$ and 40 LIBORs. It shows that the values are not stable under changes of the configuration of the numerical method. The least accurate configuration shows unrealistic spikes and negative values, indicating that the method fails to produce reasonable values in this case.

10.2 Error Analysis

The large differences in the $\tilde{L}$ in our example model appear because of the exponential functions: the centre of the Gaussian distribution is effectively shifted by the coefficient in the exponential, so that the cross-terms appearing for the earlier LIBORs are increasingly less well sampled on the grid.

The issue appears with large volatilities or a large number of LIBORs, so it will not show up in simple test cases. Effectively, this is a numerical regularisation dependent on the choice of the grid parameters. As there is no "natural" choice of grid parameters, this is not a desirable behaviour.

However, this does not necessarily affect the pricing of derivatives, as the calibration equations are still satisfied. Only if the payoff is sensitive to the ill-sampled areas of the distribution will there be an effect.

The accuracy of the interpolation scheme and the selection of grid points are tightly linked. By increasing the number of grid points the sampling can always be made arbitrarily precise, but the computational cost rises with the square of the number of points.

The error of a cubic spline approximation for a four times continuously differentiable function f is proportional to $h^4 \max |f^{(4)}|$ where h is the maximal spacing

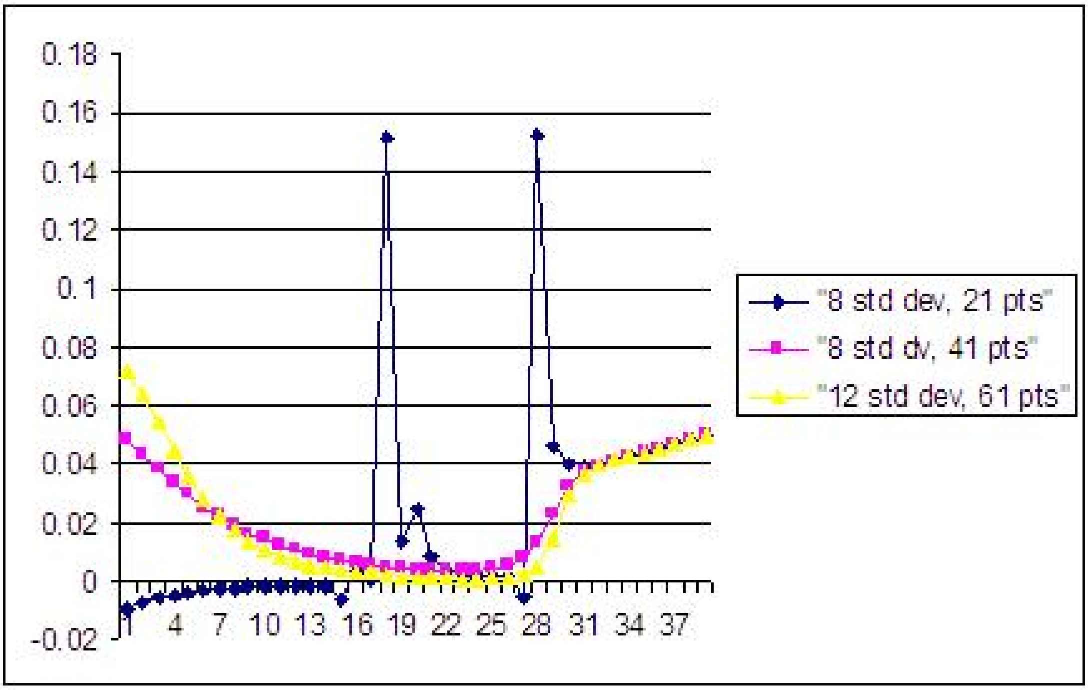

Figure 1: Example calculation of γs for various settings.

in the grid of support points of the spline. While fourth order convergence looks attractive, the fourth derivative of an exponential function can be very large and may force very small grid spacing to achieve acceptable accuracy.

10.3 Alternative Calculation Approach

We briefly examine how much further the analytical calculation we discussed above can be pushed, which should give us more precise values.

The problem with the exact representation used in the analytical calculations is that in general the number of exponential terms in the discount bond functionals grows exponentially with the number of time steps, which makes it very expensive to perform for larger numbers of time steps. The calculations we used could only be carried out efficiently because for the chosen parameters the growth of the number of factors reduced to linear.

However, we still have analytical formulae for exponentials with polynomial rather than just constant coefficients. We can choose an approximation that has only linear growth in the number of exponential terms. The idea is as follows:

- Choose a base exponential coefficient h, and allow only exponential functions of the form $x \mapsto \exp(nhx)$ where n is an integer.
- Represent an arbitrary exponential function $x \mapsto \exp(ax)$ as $p(x)\exp(nhx)$ where n is the closest integer to a/h and $p(x)$ is a Taylor expansion to a given order k of $\exp((a - nh)x)$.
- When taking products of two such terms, the resulting polynomial coefficient is again truncated at order k.
- For given k one can choose h small enough so that this approximates any exponential to a given precision for a given maximal number of multiplications.
- Calculating with terms of this type give asymptotically linear growth in the number of terms while keeping the complexity of the individual terms limited to a polynomial of order k.

This approach allows a fairly efficient calculation of the representations of the discount bond functionals to an accuracy that can be determined *a priori*. Hence it is an excellent checking tool for the standard grid implementation.

It does not extend in any straightforward way to an alternative pricing methodology, though: pricing exotic options would require taking the maximum of two functions of this type, which requires an even more complex representation of the functionals.

However, one can construct a Monte Carlo implementation from the discount bond functionals calculated with the functional representation by exponentials. This implementation is not competitive with with grid-based methods for speed but still an excellent testing tool that does not suffer from typical grid problems. The approach is also quite specific to this type of model.

The discussion illustrates a general point: specialised representations of function spaces may be efficient for particular purposes, but the general schemes used in numerical analysis are more versatile and adaptible. However, for error analysis in specific situations such specialised approaches can be very useful.

10.4 Spline Characteristics

For fixed settings, one can easily have surprises from other sources:

- The smoothing of step functions depends on the amount of variance (or effective time) in the translation operator, and if it is small (for example daily time steps in range accruals), then substantial errors may occur.

- Similarly, the Gaussian terms appearing in the analytical propagation formulae can only be well approximated if the sampling is fine enough, which again depends on the size of the time step.

Consequently, the approximation quality needs to be controlled and additional grid points introduced where necessary.

Spline interpolation has some other drawbacks; it is non-local and the resulting splines can over-oscillate relative to the function they approximate. Since we primarily interpolate with splines to apply the integration formulae, other interpolation methods that still allow use of the formulae should be explored.

Toepler [2009] analysed a number of different methods for our simple example model.

1. Interpolating with Akima interpolation or tension splines reduces problems in the sample model associated with over-fitting, and the approximation of the analytical solution of the discount bond functionals is closer for the same grid spacing.

2. The exponential behaviour can be controlled by splitting off an exponential term that captures the asymptotic behaviour and interpolating the remainder with a cubic spline. The removal of the asymptotic behaviour leads to far better spline approximation with a low number of points.

3. As the model has an analytical solution, the approximation error and hence a good grid spacing can be calculated fairly easily. For a general Markov Functional Model this is a far harder problem.

10.5 State Variables

Many products reference past history and require the introduction of state variables in a grid or tree pricing scheme, e.g. TARNs, snowballs or snowblades. Most implementations will add the state variable as an extra grid direction and apply the same scheme as before, computing values for each grid point and interpolating.

However, the interpolation relies on the smoothing of the payoff function provided by the transportation operator, and this does not hold in the direction of the state variable. In extreme cases, the function may be discontinuous in that direction (e.g. TARNs).

To control these issues, one should estimate and track the approximation error throughout the evaluation. In principle, this is not hard, as payoff functions are usually simple (if not continuous), and as analytical formulae for conditional expectations are known, it is in principle straightforward to keep track of them.

Approximation aside, the operations on a payoff function are the basic algebraic operations, addition, multiplication, and possibly division, and the creation of an indicator function. For example

$$\max(f(x), g(x)) = \mathbf{1}_{\{f-g>0\}}f(x) + \mathbf{1}_{\{g-f>0\}}g(x)$$

The error tracking scheme needs to understand the behaviour of an approximation error under these operations and compute error propagation.

Natural error measures are the L^p norms for $1 \leq p \leq \infty$. The L^∞ norm is too restrictive to use globally, as payoff functions are typically not bounded, but any other L^p norm will normally be finite. If $f, g \in L^p$, then $f + g \in L^p$ but $fg \in L^{p/2}$ for $p \geq 2$, so we cannot simply track a single L^p norm but must use a series of them, e.g. $p = 2^n$.

Over finite intervals the L^∞ norm can be used instead, so it is probably easier to track L^∞ on a grid and to use separate error measures for the residual error outside the grid. By adapting the grid dynamically to be wide enough, the function outside the grid can be neglected.

Alternatively, as demonstrated by *Toepler* [2009], one can instead keep track of the asymptotic behaviour of each functional and so reduce the size of the grid that one needs to use.

10.6 Functional Analytic Viewpoint

As an alternative to the grid as a standard numerical approach, one can take a functional analytic viewpoint. Recall that the Hermite polynomials are defined as

$$H_n(x) = (\partial^*)^n 1 = (\partial - x)^n 1 \tag{53}$$

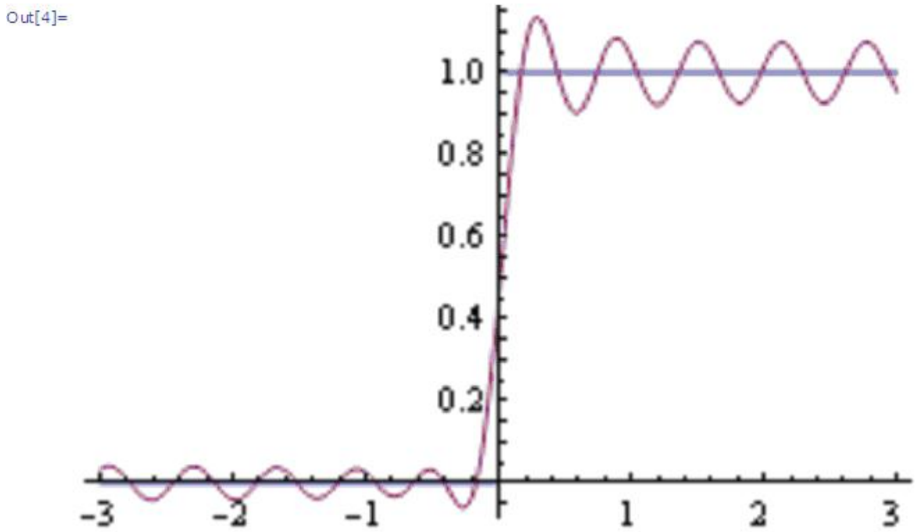

Figure 2: Approximation of step function with Hermite polynomials up to order 50.

where ∂ is the differentiation operator on $C^\infty(\mathbb{R})$ and $\partial^* = x - \partial$ is its adjoint under the Gaussian scalar product

$$\langle f, g\rangle = \int f(x)\bar{g}(x)e^{-x^2/2}\frac{dx}{\sqrt{2\pi}}.$$

The normalised Hermite polynomials

$$h_n(x) = \frac{H_n(x)}{\langle H_n, H_n\rangle^{1/2}} = \frac{H_n(x)}{\sqrt{n!}} \tag{54}$$

form an orthonormal basis of the Hilbert space of functions which have finite norm under this scalar product. They also form an eigensystem for the *Ornstein-Uhlenbeck* operator $\mathcal{L} = \partial^*\partial$ on $\mathbb{R}$.

In particular, (appropriately scaled) Hermite polynomials of order n are transported to a Hermite polynomial of order n under the transportation operator. So if a payoff function can be well approximated by a finite sum of Hermite polynomials (and as they form a Hilbert basis there is always a convergent infinite series), then we can use this as a numerical scheme.

Addition and multiplication are straightforward. However, indicator functions are difficult, as the power series for an indicator function converges very slowly. Figure 2 shows an example.

However, normalised Hermite polynomials of order n have eigenvalue n under the Ornstein-Uhlenbeck operator $\mathcal{L}$, and $\mathcal{L}$ is the infinitesimal propagator of

Brownian Motion. Hence a polynomial term of order n is multiplied by $\exp(-n\tau)$ by the transport operator over a time interval of length τ, and the contribution of high-order polynomials decays very quickly.

Therefore, the slow convergence towards the indicator function is not necessarily a problem for pricing; it depends on the problem.

10.7 A Simple Implementation

The Hermite polynomial expansion of fixed order has a very simple numerical implementation. Recall the Gauss-Hermite quadrature rules: for a given positive integer n there exist n points x_i and n weights w_i such that

$$\int_{-\infty}^{\infty} f(x)e^{-x^2}\frac{dx}{\sqrt{\pi}} = \sum_{i=1}^{n} w_i f(x_i)$$

for every polynomial f of degree $\leq 2n$. The points x_i are in fact given by the roots of the nth Hermite polynomial, and the weights can be calculated by a simple recursion formula.

By pre-calculating the values of the Hermite polynomials at the x_i this gives a simple way to calculate the expansion into Hermite polynomials up to order n of any function f.

In fact, simply representing a function by its values at the x_i transforms the calculation of the transport operator into a simple matrix multiplication.

It is difficult to make this calculation methodology robust for a general portfolio of exotic derivatives in comparison to the established grid-based methods; the instabilities introduced by the high order polynomials required to approximate discontinuities are typically more problematic than the approximations due to spline interpolation.

But at least for calibration purposes, where indicator functions do not need to be propagated except to time zero (which is a straightforward integration) this can be very powerful and fast.

Grids are clearly very flexible and robust, so they will continue to dominate in practice, but alternative methods can allow a better insight into and control of accuracy.

In summary, we can conclude the following from this discussion:

- Standard implementations of Gaussian grid pricers commonly used for Markov functional models exhibit numerical instabilities that do not show up in standard calibration or while pricing simpler derivatives.

- These instabilities can be controlled by tracking approximation errors, but in general this needs to be done dynamically in the pricing routine; a priori testing by changing grid width and spacing is not very reliable and time consuming, and needs to be repeated whenever market conditions change.

- One can make use of the information given by the propagation formulae to explicitly construct better approximations, which may not generalise to fully flexible model implementations but should allow to better track approximation errors.

11 Cross-currency Models and HJM

We now discuss how to construct models for several interest rate curves. The generic setting for this, which we consider in this section, is models for interest rate markets in several different currencies. We will see later that this approach can also cover multiple interest rate curves in the same currency.

The term structure models described in the previous section model the interest rate market for a given single currency. To model a larger market spanning several currencies, one must use a term structure model for the interest rate market of each currency and, in addition, model the exchange rates between the currencies.

In full generality, we assume that there are $K + 1$ currencies, which we will index with $\kappa = 0, \ldots, K$. Currency 0 will be used as a (domestic) base currency. We further assume that there is a fixed trading time interval $[0, T^*]$ and a filtered measure space $(\Omega, \mathcal{F})$ which captures all information flow in the markets for the given currencies.

11.1 The Risk-neutral Measure

We wish to construct a risk-neutral probability measure $\mathbb{P}$ for the market consisting of the interest rate products in all currencies as viewed by an investor using the base currency. We assume that the currencies are freely convertible so that there exist continuous adapted processes X^κ on $(\Omega, \mathcal{F})$ for $\kappa = 1, \ldots, K$ representing the spot exchange rate for currency κ into the base currency, i.e. the price of a unit of currency κ in the base currency[4]. Furthermore, we assume that for each currency κ there exists a probability measure $\mathbb{P}^\kappa$ on a sub-filtration $\mathcal{F}^\kappa$ of $\mathcal{F}$ under which the forward rates in currency κ follow risk-neutral dynamics in an HJM framework (using the money market account in currency κ, B^κ, as numeraire)

$$df^\kappa_{tT} = \sigma^\kappa_{tT} dW^\kappa_t - \sigma^\kappa_{tT} (\Sigma^\kappa_{tT})^* dt,$$

[4]Note that in an arbitrage-free model these exchange rates completely determine all exchange rates between the given currencies, so we do not need to consider other exchange rates to determine the dynamics of the model.

where W^κ denotes a Brownian motion under $\mathbb{P}^\kappa$ of dimension d_κ. Since the risk-neutral measure $\mathbb{P}$ is denominated in the base currency, we may assume without loss of generality that its restriction to $\mathcal{F}^0$ agrees with $\mathbb{P}^0$.

A tradeable claim in a foreign currency $\kappa \in \{1, \ldots, K\}$ is tradeable in the domestic market via the exchange rate. So if S_t is the price process of a tradeable in currency κ, then $X_t^\kappa S_t$ is the price process of a tradeable in the domestic currency. In an arbitrage-free model, the numeraire-adjusted price process must be a martingale. In particular, this applies to the foreign numeraire, the money market account B^κ. So the process $\frac{X_t^\kappa B_t^\kappa}{B_t^0}$ must be a continuous positive martingale in the risk-neutral measure $\mathbb{P}$ we wish to construct. As $B_0^\kappa = B_0^0 = 1$, there must exist a pre-visible process σ_t^κ on $(\Omega, \mathcal{F})$ such that

$$\frac{X_t^\kappa B_t^\kappa}{B_t^0} = X_0^\kappa \exp\left(\int_0^t \sigma_s^\kappa d\tilde{W}_s^\kappa - \frac{1}{2}\int_0^t (\sigma_s^\kappa)^2 ds\right)$$

for some Brownian motion $\tilde{W}^\kappa$ under $\mathbb{P}$, and $\int_0^t (\sigma_s^\kappa)^2 ds < \infty$ $\mathbb{P}$-almost surely. Its normalisation

$$Z_t^\kappa = \frac{X_t^\kappa B_t^\kappa}{X_0^\kappa B_t^0} = \exp\left(\int_0^t \sigma_s^\kappa d\tilde{W}_s^\kappa - \frac{1}{2}\int_0^t (\sigma_s^\kappa)^2 ds\right)$$

is the Radon-Nikodym density of the currency κ measure $\mathbb{P}^\kappa$ with respect to the risk-neutral measure $\mathbb{P}$.

By Girsanov's theorem, $\hat{W}_t^\kappa = W_t^\kappa - \int_0^t Z_s^\kappa \left\langle (Z_\cdot^\kappa)^{-1}, W_\cdot^\kappa \right\rangle_s$ is a Brownian motion under $\mathbb{P}$ and the currency κ forward rates have the dynamics

$$\begin{aligned} f_{tT}^\kappa &= f_{0T}^\kappa + \int_0^t \sigma_{sT}^\kappa d\hat{W}_s^\kappa - \int_0^t \sigma_{sT}^\kappa (\Sigma_{sT}^\kappa)^* ds - \int_0^t \frac{1}{Z_s^\kappa} d\langle Z_\cdot^\kappa, f_{\cdot T}\rangle_s \\ &= f_{0T}^\kappa + \int_0^t \sigma_{sT}^\kappa d\hat{W}_s^\kappa - \int_0^t \sigma_{sT}^\kappa \left((\Sigma_{sT}^\kappa)^* + \rho_s^\kappa \sigma_s^\kappa\right) ds \end{aligned} \tag{55}$$

where ρ_s^κ is the vector of instantaneous correlations between the Brownian motions $\tilde{W}_s^\kappa$ and $\hat{W}_s^{\kappa,i}$ at time s, i.e. $\rho_s^{\kappa,i} ds = d\left\langle \tilde{W}_s^\kappa, \hat{W}_s^{\kappa,i}\right\rangle$.

Analogously, the dynamics of the foreign discount bonds discounted by the corresponding money market account under $\mathbb{P}$ is

$$\begin{aligned} \frac{P_{tT}^\kappa}{B_t^\kappa} &= P_{0T}^\kappa \exp\left(\int_0^t \Sigma_{sT}^\kappa d\hat{W}_s^\kappa - \frac{1}{2}\int_0^t \Sigma_{sT}^\kappa (\Sigma_{sT}^\kappa)^* ds\right) - \int_0^t \frac{1}{Z_s^\kappa} d\left\langle Z_\cdot^\kappa, \frac{P_{\cdot T}^\kappa}{B_\cdot^\kappa}\right\rangle_s \\ &= P_{0T}^\kappa \exp\left(\int_0^t \Sigma_{sT}^\kappa d\hat{W}_s^\kappa - \int_0^t \Sigma_{sT}^\kappa \left(\frac{1}{2}(\Sigma_{sT}^\kappa)^* + \rho_s^\kappa \sigma_s^\kappa\right) ds\right). \end{aligned} \tag{56}$$

The price process of a foreign currency discount bond expiring at time T is $X_t^\kappa P_{tT}^\kappa$. Discounted by the domestic numeraire it is a martingale, the process of which is

$$
\begin{aligned}
\frac{X_t^\kappa P_{tT}^\kappa}{B_t^0} = X_0^\kappa P_{0T}^\kappa \exp\Bigg(& \int_0^t \Sigma_{sT}^\kappa d\hat{W}_s^\kappa + \int_0^t \sigma_s^\kappa d\tilde{W}_s^\kappa \\
& -\frac{1}{2}\int_0^t \left(\Sigma_{sT}^\kappa (\Sigma_{sT}^\kappa)^* + 2\Sigma_{sT}^\kappa \rho_s^\kappa \sigma_s^\kappa + (\sigma_s^\kappa)^2\right) ds\Bigg) \qquad (57)
\end{aligned}
$$

Note that although we defined σ_t^κ as the volatility of a cash investment in the currency κ money market account by a domestic investor, it is also the volatility of the foreign exchange rate X_t^κ, since

$$\frac{dZ_t^\kappa}{Z_t^\kappa} = \frac{dX_t^\kappa}{X_t^\kappa} + \frac{dB_t^\kappa}{B_t^\kappa} - \frac{dB_t^0}{B_t^0} + (\dots)dt$$

and $dB_t^\kappa = B_t^\kappa r_t dt$ is of finite variation for all κ.

It is often assumed that the return on a foreign currency follows a log-normal process, i.e. that σ_t^κ is a deterministic function of t, but far more general dynamics can be accommodated in this framework by choosing an appropriate pre-visible process on $(\Omega, \mathcal{F})$.

11.2 Single Driving Brownian Motion

The correlated Brownian motions W^0, $\hat{W}^\kappa$ and $\tilde{W}^\kappa$ for $\kappa = 1, \dots, K$ under $\mathbb{P}$ can be replaced by a single (multi-dimensional) Brownian motion. In fact, there exist a Brownian motion W, of some dimension d, and functions $\hat{\rho}^\kappa : [0, T^*] \to L(\mathbb{R}^d, \mathbb{R}^{d_\kappa})$ for $\kappa = 0, \dots, K$ and $\tilde{\rho}^\kappa : [0, T^*] \to L(\mathbb{R}^d, \mathbb{R})$ for $\kappa = 1, \dots, K$ such that $\hat{W}_t^\kappa = \int_0^t \hat{\rho}_s^\kappa dW_s$ and $\tilde{W}_t^\kappa = \int_0^t \tilde{\rho}_s^\kappa dW_s$. W may be chosen to extend W^0 to d dimensions, so that $\hat{\rho}_t^0$ is the projection onto the first d_κ coordinates for all t. Using W, (55) may be written as

$$f_{tT}^\kappa = f_{0T}^\kappa + \int_0^t \sigma_{sT}^\kappa \hat{\rho}_s^\kappa dW_s - \int_0^t \sigma_{sT}^\kappa \hat{\rho}_s^\kappa \left((\Sigma_{sT}^\kappa \hat{\rho}_s^\kappa)^* + (\tilde{\rho}_s^\kappa)^* \sigma_s^\kappa\right) ds, \qquad (58)$$

and the domestic forward rates can be decomposed as

$$f_{tT}^0 = f_{0T}^0 + \int_0^t \sigma_{sT}^0 \hat{\rho}_s^0 dW_s - \int_0^t \sigma_{sT}^0 \hat{\rho}_s^0 (\Sigma_{sT}^0 \hat{\rho}_s^0)^* ds. \qquad (59)$$

Equation (57) can be written as

$$\frac{X_t^\kappa P_{tT}^\kappa}{B_t^0} = X_0^\kappa P_{0T}^\kappa \exp\left(\int_0^t (\Sigma_{sT}^\kappa \hat{\rho}_s^\kappa + \sigma_s^\kappa \tilde{\rho}_s^\kappa) dW_s - \frac{1}{2}\int_0^t \|\Sigma_{sT}^\kappa \hat{\rho}_s^\kappa + \sigma_s^\kappa \tilde{\rho}_s^\kappa\|^2 ds\right) \qquad (60)$$

while equation (56) can be written as

$$\frac{P^\kappa_{tT}}{B^\kappa_t} = P^\kappa_{0T} \exp\left(\int_0^t \Sigma^\kappa_{sT} \hat{\rho}^\kappa_s dW_s - \int_0^t \Sigma^\kappa_{sT} \hat{\rho}^\kappa_s \left(\frac{1}{2} \Sigma^\kappa_{sT} \hat{\rho}^\kappa_s + \sigma^\kappa_s \tilde{\rho}^\kappa_s \right)^* ds \right). \tag{61}$$

It follows that replacing σ^κ_{tT} by $\sigma^\kappa_{tT}\hat{\rho}^\kappa_t$ we could assume that the measures $\mathbb{P}^0$ and $\mathbb{P}$ are identical, that the measures $\mathbb{P}^\kappa$ for $\kappa \neq 0$ are equivalent to $\mathbb{P}$ and that the Brownian motions W^κ are the martingale parts of W under $\mathbb{P}^\kappa$. Furthermore, replacing σ^κ_t by $\sigma^\kappa_t \tilde{\rho}^\kappa_t$ we could assume that $\tilde{W}^\kappa$ is identical to W for all κ. This simplifies formulae and the manipulation of the model, but can be inconvenient if the models for the individual currencies are exogenously specified.

To avoid confusion, we will not make this assumption in the following, but only define

$$\sigma_{tT} = \sigma^0_{tT} \hat{\rho}^0_t \qquad \text{and} \qquad \Sigma_{tT} = \Sigma^0_{tT} \hat{\rho}^0_t \tag{62}$$

to be able to drop the superscript 0 for the domestic volatilities.

11.3 Forward Measures

As for the single currency models, the time T forward measure $\mathbb{P}_T$ is obtained by changing numeraire to the domestic discount bond expiring at time T. From equations (7) and (60) the dynamics of a foreign discount bond expiring at time S under this measure is

$$\begin{aligned} \frac{X^\kappa_t P^\kappa_{tS}}{P^0_{tT}} = \frac{X^\kappa_t P^\kappa_{tS}}{B^0_t} \left(\frac{P^0_{tT}}{B^0_t} \right)^{-1} = \frac{X^\kappa_0 P^\kappa_{0S}}{P^0_{0T}} \exp\bigg(& \int_0^t (\Sigma^\kappa_{sS} \hat{\rho}^\kappa_s + \sigma^\kappa_s \tilde{\rho}^\kappa_s - \Sigma_{sT}) dW^T_s \\ & - \frac{1}{2} \|\Sigma^\kappa_{sS} \hat{\rho}^\kappa_s + \sigma^\kappa_s \tilde{\rho}^\kappa_s - \Sigma_{sT}\|^2 ds \bigg) \end{aligned} \tag{63}$$

where W^T is again a Brownian motion under $\mathbb{P}_T$.

Alternatively, one may consider the *forward exchange rate* for time T,

$$X^{\kappa,T}_t = \frac{X^\kappa_t P^\kappa_{tT}}{P^0_{tT}} = \frac{X^\kappa_0 P^\kappa_{0T}}{P^0_{0T}} \exp\left(\int_0^t \sigma^{\kappa,T}_s dW_s + \int_0^t \sigma^{\kappa,T}_s \left(\Sigma_{tT} - \frac{1}{2} \sigma^{\kappa,T}_s \right)^* ds \right) \tag{64}$$

where

$$\sigma^{\kappa,T}_t = \sigma^\kappa_t \tilde{\rho}^\kappa_t + \Sigma^\kappa_{tT} \hat{\rho}^\kappa_t - \Sigma_{tT} \tag{65}$$

is the time T forward volatility for currency κ. Note that the forward volatility depends on the chosen forward expiry T; the forward volatilities for two different expiries T and S are related by

$$\sigma^{\kappa,S}_t = \sigma^{\kappa,T}_t + (\Sigma^\kappa_{tS} - \Sigma^\kappa_{tT}) \hat{\rho}^\kappa_t - (\Sigma_{tS} - \Sigma_{tT}). \tag{66}$$

The normalised forward exchange rate $\frac{X_t^{\kappa,T}}{X_0^{\kappa,T}}$ is the Radon-Nikodym density of the currency κ forward measure $\mathbb{P}_T^\kappa$ with respect to the forward measure $\mathbb{P}_T$, and under $\mathbb{P}_T$ the forward exchange rate is a martingale described by the equation

$$X_t^{\kappa,T} = X_0^{\kappa,T} \exp\left(\int_0^t \sigma_s^{\kappa,T} dW_s^T - \frac{1}{2}\int_0^t \sigma_s^{\kappa,T}\left(\sigma_s^{\kappa,T}\right)^* ds\right). \tag{67}$$

It follows that (63) can be re-written as

$$\frac{X_t^\kappa P_{tS}^\kappa}{P_{tT}^0} = \frac{X_t^{\kappa,T} P_{tS}^\kappa}{P_{tT}^\kappa} = \frac{X_0^{\kappa,T} P_{0S}^\kappa}{P_{0T}^\kappa} \exp\left(\int_0^t \left[(\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa)\hat{\rho}_s^\kappa + \sigma_s^{\kappa,T}\right] dW_s^T \right.$$
$$\left. -\frac{1}{2}\int_0^t \|(\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa)\hat{\rho}_s^\kappa + \sigma_s^{\kappa,T}\|^2 ds\right) \tag{68}$$

which only references the dynamics under the time T forward measures $\mathbb{P}_T$ and $\mathbb{P}_T^\kappa$. Similarly, one can write

$$\frac{P_{tS}^\kappa}{P_{tT}^\kappa} = \frac{P_{0S}^\kappa}{P_{0T}^\kappa} \exp\left(\int_0^t (\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa)\hat{\rho}_s^\kappa dW_s^T \right.$$
$$\left. - \int_0^t (\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa)\hat{\rho}_s^\kappa \left(\frac{1}{2}(\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa)\hat{\rho}_s^\kappa + \sigma_s^{\kappa,T}\right)^* ds\right). \tag{69}$$

A non-degenerate forward exchange rate volatility function $\sigma^{\kappa,T}$ can be split into its total volatility $\tilde{\sigma}_t^{\kappa,T} = \|\sigma_t^{\kappa,T}\|$ and a unit correlation vector $\tilde{\rho}_t^{\kappa,T} = \frac{\sigma_t^{\kappa,T}}{\tilde{\sigma}_t^{\kappa,T}}$ so that the forward exchange rate is

$$X_t^{\kappa,T} = X_0^{\kappa,T} \exp\left(\int_0^t \tilde{\sigma}_s^{\kappa,T} d\tilde{W}_s^{\kappa,T} - \frac{1}{2}\int_0^t (\tilde{\sigma}_s^{\kappa,T})^2 ds\right) \tag{70}$$

where $\tilde{W}^{\kappa,T}$ is the Brownian motion under the forward measure $\mathbb{P}_T$ defined by $d\tilde{W}_t^{\kappa,T} = \tilde{\rho}_t^{\kappa,T} dW_t^T$. Using this, the dynamics for the numeraire-adjusted foreign discount bonds under $\mathbb{P}_T$ in (68) can be re-written as

$$\frac{X_t^{\kappa,T} P_{tS}^\kappa}{P_{tT}^\kappa} = \frac{X_0^{\kappa,T} P_{0S}^\kappa}{P_{0T}^\kappa} \exp\left(\int_0^t (\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa) d\hat{W}_s^{\kappa,T} + \int_0^t \tilde{\sigma}_s^{\kappa,T} d\tilde{W}_s^{\kappa,T} \right.$$
$$\left. -\frac{1}{2}\int_0^t \left[(\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa)^2 + 2(\Sigma_{sS}^\kappa - \Sigma_{sT}^\kappa)\rho_s^{\kappa,T}\tilde{\sigma}_s^{\kappa,T} + (\tilde{\sigma}_s^{\kappa,T})^2\right] ds\right) \tag{71}$$

where $\rho_t^{\kappa,T} = \hat{\rho}_t^\kappa (\tilde{\rho}_t^{\kappa,T})^*$ is the instantaneous correlation between the Brownian motions $\hat{W}^{\kappa,T}$ and $\tilde{W}^{\kappa,T}$, which drive the currency κ forward rates and the time T forward exchange rate $X^{\kappa,T}$ respectively. This format is useful as calibration procedures often specify the dynamics of the interest rates in each currency and of the (forward) exchange rates separately, and then combine the resulting processes by specifying the correlation between their driving Brownian motions.

12 Multi-currency (Generalised) Vasicek Models

As the generalised Vasicek model is a Gaussian HJM model, the above specialises in a straightforward manner for this class of model. We derive explicit formulae for the case that the interest rate curve of every currency is modelled by a single-factor generalised Vasicek model. We the comment briefly on the generalisation to multi-factor generalised Vasicek models.

12.1 Multi-currency With Single-factor Models

Assume that we have a multi-currency model in which the interest rate curve of each currency κ is modelled by a one-factor generalised Vasicek model

$$dr_t^\kappa = (a^\kappa(t) - b^\kappa(t) r_t^\kappa)dt + \sigma^\kappa(t) dW_t^\kappa.$$

Defining $\phi_t^\kappa = \int_0^t \exp\left(-\int_0^u b^\kappa(s)ds\right) du$ and $g_t^\kappa = \exp\left(\int_0^t b^\kappa(s)ds\right)\sigma^\kappa(t)$, it follows from (22), (57) and (60) that we have in the risk-neutral measure

$$\frac{P_{tT}^0}{B_t^0} = P_{0T}^0 \exp\left(-\int_0^t (\phi_T^0 - \phi_u^0) g_u^0 dW_u - \frac{1}{2}\int_0^t (\phi_T - \phi_u)^2 g_u^2 du\right) \tag{72}$$

for the domestic discount bonds and

$$\begin{aligned}\frac{X_t^\kappa P_{tT}^\kappa}{B_t^0} = X_0^\kappa P_{0T}^\kappa \exp\Bigg(&-\int_0^t (\phi_T^\kappa - \phi_u^\kappa) g_u^\kappa d\hat{W}_u^\kappa + \int_0^t \sigma_u^\kappa d\tilde{W}_u^\kappa \\ &-\frac{1}{2}\int_0^t \left[(\phi_T^\kappa - \phi_u^\kappa)^2 (g_u^\kappa)^2 - 2(\phi_T^\kappa - \phi_u^\kappa) g_u^\kappa \sigma_u^\kappa \rho_u^\kappa + (\sigma_u^\kappa)^2\right] du\Bigg)\end{aligned} \tag{73}$$

for the foreign currency κ discount bonds, where ρ_t^κ is the instantaneous correlation between the spot exchange rate of currency κ and the currency κ short rate. In terms of a single driving Brownian motion we can write (73) as

$$\begin{aligned}\frac{X_t^\kappa P_{tT}^\kappa}{B_t^0} = X_0^\kappa P_{0T}^\kappa \exp\Bigg(&\int_0^t \left(-(\phi_T^\kappa - \phi_u^\kappa) g_u \hat{\rho}_u^\kappa + \sigma_u^\kappa \tilde{\rho}_u^\kappa\right) dW_u \\ &-\frac{1}{2}\int_0^t \| - (\phi_T^\kappa - \phi_u^\kappa) g_u \hat{\rho}_u^\kappa + \sigma_u^\kappa \tilde{\rho}_u^\kappa \|^2 du\Bigg).\end{aligned} \tag{74}$$

Similarly, we have

$$\begin{aligned}\frac{P_{tT}^\kappa}{B_t^\kappa} = P_{0T}^\kappa \exp\Bigg(&-\int_0^t (\phi_T^\kappa - \phi_u^\kappa) g_u^\kappa d\hat{W}_u^\kappa \\ &-\int_0^t \left[\frac{1}{2}(\phi_T^\kappa - \phi_u^\kappa)^2 (g_u^\kappa)^2 - (\phi_T^\kappa - \phi_u^\kappa) g_u^\kappa \sigma_u^\kappa \rho_u^\kappa\right] du\Bigg)\end{aligned} \tag{75}$$

and

$$\frac{P^{\kappa}_{tT}}{B^{\kappa}_{t}} = P^{\kappa}_{0T}\exp\Bigg(-\int_0^t (\phi^{\kappa}_{T}-\phi^{\kappa}_{u})g_u\hat{\rho}^{\kappa}_{u}dW_u$$
$$-\int_0^t\left[\frac{1}{2}\left\|(\phi^{\kappa}_{T}-\phi^{\kappa}_{u})g_u\hat{\rho}^{\kappa}_{u}\right\|-(\phi^{\kappa}_{T}-\phi^{\kappa}_{u})g_u\hat{\rho}^{\kappa}_{u}(\sigma^{\kappa}_{u}\bar{\rho}^{\kappa}_{u})^{*}\right]du\Bigg). \quad (76)$$

12.2 Forward Measures

In the time T forward measure, we obtain, as before in the single-currency case,

$$\frac{P^{0}_{tS}}{P^{0}_{tT}} = \frac{P^{0}_{0S}}{P^{0}_{0T}}\exp\left(-(\phi^{0}_{S}-\phi^{0}_{T})\int_0^t g^{0}_{u}dW^{T}_{u}-\frac{1}{2}(\phi^{0}_{S}-\phi^{0}_{T})^2\int_0^t (g^{0}_{u})^2du\right) \quad (77)$$

for the domestic discount bond expiring at time S, and

$$\frac{X^{\kappa,T}_{t}P^{\kappa}_{tS}}{P^{\kappa}_{tT}} = \frac{X^{\kappa,T}_{t}P^{\kappa}_{0S}}{P^{\kappa}_{0T}}\exp\Bigg(\int_0^t\left(-(\phi^{\kappa}_{S}-\phi^{\kappa}_{T})g^{\kappa}_{u}\hat{\rho}^{\kappa}_{u}+\sigma^{\kappa,T}_{u}\right)dW^{T}_{u}$$
$$-\frac{1}{2}\int_0^t \|-(\phi^{\kappa}_{S}-\phi^{\kappa}_{T})g^{\kappa}_{u}\hat{\rho}^{\kappa}_{u}+\sigma^{\kappa,T}_{u}\|^2du\Bigg) \quad (78)$$

for the foreign currency κ discount bond expiring at time S.

As in (71) we can re-write this as

$$\frac{X^{\kappa,T}_{t}P^{\kappa}_{tS}}{P^{\kappa}_{tT}} = \frac{X^{\kappa,T}_{0}P^{\kappa}_{0S}}{P^{\kappa}_{0T}}\exp\Bigg(-(\phi^{\kappa}_{S}-\phi^{\kappa}_{T})\int_0^t g^{\kappa}_{u}d\hat{W}^{\kappa,T}_{u}+\int_0^t\tilde{\sigma}^{\kappa,T}_{u}d\tilde{W}^{\kappa,T}_{u}$$
$$-\frac{1}{2}\int_0^t\left[(\phi^{\kappa}_{S}-\phi^{\kappa}_{T})^2(g^{\kappa}_{u})^2-2(\phi^{\kappa}_{S}-\phi^{\kappa}_{T})g^{\kappa}_{u}\rho^{\kappa,T}_{u}\tilde{\sigma}^{\kappa,T}_{u}+(\tilde{\sigma}^{\kappa,T}_{u})^2\right]du\Bigg) \quad (79)$$

where $\rho^{\kappa,T}_{t}$ is again the instantaneous correlation between the Brownian motions $\hat{W}^{\kappa,T}$ driving the currency κ forward rates and $\tilde{W}^{\kappa,T}$ driving the forward exchange rate $X^{\kappa,T}$ under the time T forward measure $\mathbb{P}_T$.

Similarly we have

$$\frac{P^{\kappa}_{sS}}{P^{\kappa}_{sT}} = \frac{P^{\kappa}_{0S}}{P^{\kappa}_{0T}}\exp\Bigg(\int_0^t(\phi^{\kappa}_{T}-\phi^{\kappa}_{S})g^{\kappa}_{u}d\hat{W}^{\kappa,T}_{u}-\int_0^t\frac{1}{2}(\phi^{\kappa}_{S}-\phi^{\kappa}_{T})^2(g^{\kappa}_{u})^2du$$
$$+\int_0^t(\phi^{\kappa}_{S}-\phi^{\kappa}_{T})g^{\kappa}_{u}\rho^{\kappa,T}_{u}\tilde{\sigma}^{\kappa,T}du\Bigg). \quad (80)$$

12.3 Using the ϕ/G Parametrisation

Using the above notation, one defines $G_t^\kappa = \int_0^t (g_u^\kappa)^2 du$ for all κ as well as $G_t^{\kappa,T} = \int_0^t (\tilde{\sigma}_u^{\kappa,T})^2 du$ and $G_t^{\kappa\kappa,T} = \int_0^t g_u^\kappa \tilde{\sigma}_u^{\kappa,T} \rho_u^{\kappa,T} du$ for all $\kappa \geq 1$. With these functions, one can write the dynamics of the numeraire-adjusted discount bonds under the time T forward measure as

$$\frac{P_{tS}^0}{P_{tT}^0} = \frac{P_{0S}^0}{P_{0T}^0} \exp\left(-(\phi_S^0 - \phi_T^0)\check{W}_{G_t^0}^{0,T} - \frac{1}{2}(\phi_S^0 - \phi_T^0)^2 G_t^0\right) \tag{81}$$

for the domestic bonds, following (77), and

$$\begin{aligned}\frac{X_t^{\kappa,T} P_{tS}^\kappa}{P_{tT}^\kappa} = \frac{X_0^{\kappa,T} P_{0S}^\kappa}{P_{0T}^\kappa} \exp\Big(&-(\phi_S^\kappa - \phi_T^\kappa)\check{W}_{G_t^\kappa}^{\kappa,T} + \bar{W}_{G_t^{\kappa,T}}^{\kappa,T} \\ &-\frac{1}{2}(\phi_S^\kappa - \phi_T^\kappa)^2 G_t^\kappa + (\phi_S^\kappa - \phi_T^\kappa) G_t^{\kappa\kappa,T} - \frac{1}{2} G_t^{\kappa,T}\Big)\end{aligned} \tag{82}$$

for the foreign currency κ bonds, following (79), where $\check{W}^{\kappa,T}$ is a time-changed version of $\hat{W}^{\kappa,T}$ and $\bar{W}^{\kappa,T}$ a time-changed version of $\tilde{W}^{\kappa,T}$. Analogously, we obtain

$$\frac{P_{sS}^\kappa}{P_{sT}^\kappa} = \frac{P_{0S}^\kappa}{P_{0T}^\kappa} \exp\left(-(\phi_S^\kappa - \phi_T^\kappa)\check{W}_{G_t^\kappa}^{\kappa,T} - \frac{1}{2}(\phi_S^\kappa - \phi_T^\kappa)^2 G_t^\kappa + (\phi_S^\kappa - \phi_T^\kappa) G_t^{\kappa\kappa,T}\right). \tag{83}$$

For a full specification of the model, one must also specify the remaining covariance functions between the time-changed Brownian motions $\check{W}^{\kappa,T}$ and $\bar{W}^{\kappa,T}$,

$$G_t^{\kappa\lambda} = \int_0^t g_u^\kappa g_u^\lambda d\left\langle \hat{W}^\kappa, \hat{W}^\lambda \right\rangle_u = \int_0^t g_u^\kappa g_u^\lambda \hat{\rho}_u^\kappa (\hat{\rho}_u^\lambda)^* du, \tag{84}$$

$$G_t^{\kappa\lambda,T} = \int_0^t g_u^\kappa \tilde{\sigma}_u^{\lambda,T} d\left\langle \hat{W}^\kappa, \tilde{W}^{\lambda,T} \right\rangle_u = \int_0^t g_u^\kappa \tilde{\sigma}_u^{\lambda,T} \hat{\rho}_u^\kappa (\rho_u^{\kappa,T})^* du \tag{85}$$

and

$$\tilde{G}_t^{\kappa\lambda,T} = \int_0^t \tilde{\sigma}_u^{\kappa,T} \tilde{\sigma}_u^{\lambda,T} d\left\langle \tilde{W}^{\kappa,T}, \tilde{W}^{\lambda,T} \right\rangle_u = \int_0^t \tilde{\sigma}_u^{\kappa,T} \tilde{\sigma}_u^{\lambda,T} \rho_u^{\kappa,T} (\rho_u^{\lambda,T})^* du. \tag{86}$$

It is important to note that while the G^κ and $G^{\kappa\lambda}$ do not depend on the chosen forward measure, we use a different forward exchange rate for every forward measure, so the $G^{\kappa,T}$, $G^{\kappa\lambda,T}$ and $\tilde{G}^{\kappa\lambda,T}$ change with the choice of forward measure $\mathbb{P}_T$.

For a change from the time T to the time T' forward measure, we obtain from (66)

$$
\begin{aligned}
\tilde{G}_t^{\kappa\lambda,T'} &= \int_0^t \sigma_u^{\kappa,T'}(\sigma_u^{\lambda,T'})^* du \\
&= \int_0^t \left(\sigma_u^{\kappa,T} + (\phi_{T'} - \phi_T) g_u \hat{\rho}_u^0 - (\phi_{T'}^\kappa - \phi_T^\kappa) g_u^\kappa \hat{\rho}_u^\kappa\right) \\
&\quad \times \left(\sigma_u^{\lambda,T} + (\phi_{T'} - \phi_T) g_u \hat{\rho}_u^0 - (\phi_{T'}^\lambda - \phi_T^\lambda) g_u^\lambda \hat{\rho}_u^\lambda\right)^* du \\
&= \tilde{G}^{\kappa\lambda,T} + (\phi_{T'} - \phi_T)(G_t^{0\kappa,T} + G_t^{0\lambda,T}) + (\phi_{T'} - \phi_T)^2 G_t^0 \\
&\quad - (\phi_{T'}^\kappa - \phi_T^\kappa) G_t^{\kappa\lambda,T} - (\phi_{T'}^\lambda - \phi_T^\lambda) G_t^{\lambda\kappa,T} \\
&\quad - (\phi_{T'} - \phi_T)\left[(\phi_{T'}^\kappa - \phi_T^\kappa) G_t^{0\kappa} + (\phi_{T'}^\lambda - \phi_T^\lambda) G_t^{0\lambda}\right] \\
&\quad + (\phi_{T'}^\kappa - \phi_T^\kappa)(\phi_{T'}^\lambda - \phi_T^\lambda) G_t^{\kappa\lambda}
\end{aligned} \tag{87}
$$

for the covariances between the Brownian motions driving the exchange rates for two currencies κ and λ. In particular, we have for the variance of one of these Brownian motions

$$
\begin{aligned}
G_t^{\kappa,T'} &= G^{\kappa,T} + 2(\phi_{T'} - \phi_T) G_t^{0\kappa,T} + (\phi_{T'} - \phi_T)^2 G_t^0 - 2(\phi_{T'}^\kappa - \phi_T^\kappa) G_t^{\kappa\kappa,T} \\
&\quad - 2(\phi_{T'} - \phi_T)(\phi_{T'}^\kappa - \phi_T^\kappa) G_t^{0\kappa} + (\phi_{T'}^\kappa - \phi_T^\kappa)^2 G_t^\kappa.
\end{aligned} \tag{88}
$$

For the covariance of the Brownian motion driving of the currency κ yield curve with the Brownian motion driving the currency λ exchange rate, we obtain

$$
\begin{aligned}
G_t^{\kappa\lambda,T'} &= \int_0^t g_u^\kappa \hat{\rho}_u^\kappa (\sigma_u^{\lambda,T'})^* du \\
&= \int_0^t g_u^\kappa \hat{\rho}_u^\kappa \left(\sigma_u^{\lambda,T} + \hat{\rho}_u^\lambda g_u^\lambda (\phi_T^\lambda - \phi_{T'}^\lambda) - \hat{\rho}_u^0 g_u (\phi_T - \phi_{T'})\right)^* du \\
&= \int_0^t g_u^\kappa \hat{\rho}_u^\kappa (\sigma^{\lambda,T})^* du + (\phi_T^\lambda - \phi_{T'}^\lambda) \int_0^t g_u^\kappa g_u^\lambda du \\
&\quad - (\phi_T - \phi_{T'}) \int_0^t g_u^\kappa g_u \hat{\rho}_u^\kappa (\hat{\rho}_u^0)^* du \\
&= G_t^{\kappa\lambda,T} + (\phi_T^\lambda - \phi_{T'}^\lambda) G_t^{\kappa\lambda} - (\phi_T - \phi_{T'}) G_t^{0\kappa}.
\end{aligned} \tag{89}
$$

12.4 Multi-currency With Multi-factor Models

The above generalises easily to the multi-factor generalised Vasicek models we described in 5.3. They can be developed analogously to the above, so we only give the basics here.

Let $g_t^{\kappa,i}$, $\phi_t^{\kappa,i}$ and $\rho_t^{\kappa,i}$ be the parameter functions describing the dynamics of the generalised Vasicek model for currency κ, which we assume to have n_κ factors $W^{\kappa,i}$. Define $G_t^{\kappa,ij} = \int_0^t g_u^{\kappa,i} g_u^{\kappa,j} \rho_u^{\kappa,ij} du$. In addition let, as above, $\tilde{\sigma}_t^{\kappa,T}$ be the volatility of the time T forward exchange rate $X_t^{\kappa,T}$ for currency κ and denote the correlation between its driving Brownian motion and $W^{\kappa,i}$ by $\rho_t^{\kappa,i}$.

Then the dynamics of the domestic discount bonds are as in (39) or (40), and the foreign currency κ discount bonds follow

$$\frac{X_t^{\kappa,T} P_{tS}^\kappa}{P_{tT}^\kappa} = \frac{X_0^{\kappa,T} P_{0S}^\kappa}{P_{0T}^\kappa} \exp\Bigg(-\sum_{i=1}^{n_\kappa} (\phi_S^{\kappa,i} - \phi_T^{\kappa,i}) \int_0^t g_u^{\kappa,i} d\hat{W}_u^{\kappa,T,i} + \int_0^t \tilde{\sigma}_u^{\kappa,T} d\tilde{W}_u^{\kappa,T}$$
$$-\frac{1}{2} \sum_{i,j=1}^{n_\kappa} (\phi_S^{\kappa,i} - \phi_T^{\kappa,i})(\phi_S^{\kappa,j} - \phi_T^{\kappa,j}) G_t^{\kappa,ij} + \sum_{i=1}^{n_\kappa} (\phi_S^{\kappa,i} - \phi_T^{\kappa,i}) \bar{G}_t^{\kappa,i} - \frac{1}{2} G_t^{\kappa,T} \Bigg) \quad (90)$$

where $G_t^{\kappa,T} = \int_0^t (\bar{\sigma}_u^{\kappa,T})^2 du$, $\bar{G}_t^{\kappa,i} = \int_0^t g_u^{\kappa,i} \tilde{\sigma}_u^{\kappa,T} \rho_u^{\kappa,i} du$, and the $\hat{W}^{\kappa,T,i}$ are Brownian motions under the time T forward measure with instantaneous correlations $\rho_t^{\kappa,ij}$.

12.5 Application to Quantos

One straightforward application is the pricing of quanto trades. Consider any interest rate derivative in a currency $\kappa = 1$ and consider the modification under which the payout is in another currency $\kappa = 0$ at some fixed exchange rate (usually 1). While the value of a cash payment in currency $\kappa = 1$ is a martingale in currency $\kappa = 0$ when converted at market exchange rates, the *numerical* value of such a payment is not a martingale. Instead, the SDE of the discount bond has a drift term.

In a generalised Vasicek model this has a very simple deterministic form, as in (75), (76), (80) or (83). One simply needs to multiply all discount bonds with a deterministic factor to adjust the dynamics of a generalised Vasicek model to a numeraire in another currency. It is therefore unnecessary to explicitly model the exchange rate process in the valuation of quantos, and in some cases the dynamics of the interest rates in the other currency also do not need to be explicitly modelled. For example, quanto caplet in a multi-currency generalised Vasicek model can be valued with the same pricing formula, except that the forward needs to be adjusted (quanto adjustment).

For general Gaussian HJM models the dynamics of discount bonds similarly change only by deterministic (exponential) factors, but these are considerably more complicated to compute.

13 Multiple Curves in a Single Market

We now discuss how to handle multiple curves in a single interest rate market, e.g. an OIS curve for discounting and LIBOR curves with various frequencies for reference rates, as has become necessary since the Global Financial Crisis of 2008/09. Since we discussed the background and need for this in some detail in Module 5, we discuss only the modelling aspects here. As we indicated above, the techniques for multi-currency models can be applied. However, this requires justification, since there is no obvious FX rate that decouples the different curves. The differences between the curves arises from credit (and liquidity) risk, which needs to be represented appropriately.

13.1 Basic Assumptions

Let us assume that we have an agreed market standard on the marginal "safe" cost of funding, and that we can therefore agree between market participants on the value of future cash flows at any point in time (subject to some standard collateralisation or margining agreement). We denote by $P(t,T)$ the time t value of a cash flow at time T. At a fixed point in time t the function $T \mapsto P(t,T)$ is then the discount curve. We can define instantaneously compounded forward rates $f_{tT} = -\frac{\partial \log P(t,T)}{\partial T}$.

For example, in the USD market $T \mapsto P(t,T)$ should be the OIS discount curve. As standard market quotes for overnight index swaps are collateralised and hence discount rates and forward rates are consistent for OIS instruments, any standard curve construction can be used to construct an OIS curve.

For a market reference index rate I in the same currency, for example three month USD BBA LIBOR, it is not true in general that spot or forward rates for that index can be implied from the discount curve. However, as the index rate gives an average cost of funding at a certain tenor for a particular group of institutions, it is a reasonable assumption that at any point in time we can define a discount curve that corresponds to the cost of funding at this rate. We therefore assume that at any time t we can define a discount function $T \mapsto P^I(t,T)$ such that the forward index rate for time T is

$$I(T;t) = \frac{1}{\alpha_{I,T}}\left(\frac{P^I(t,T)}{P^I(t,m_I(T))} - 1\right) \tag{91}$$

where $\alpha_{I,T}$ is the day count fraction and $m_I(T)$ the maturity date of the forward rate. As P^I is a discount curve we have $P^I(t,t) = 1$, and $P^I(t,T)$ can be interpreted as the discount factor for a future cash flow at T from a rolling investment at the index rate. We call $T \mapsto P^I(t,T)$ the *index curve* or *prediction curve* for index I at time t.

The agreed discount rate in the market should be almost risk free, so generally we would expect the index curve to have a spread over the funding curve, i.e. defining the index forwards as $f^I(t,T) \equiv f^I_{tT} = -\frac{\partial \log P^I(t,T)}{\partial T}$ we expect $f^I_{tT} = f_{tT} + s^I_{tT}$ for some positive function $T \mapsto s^I_{tT} \equiv s^I(t,T)$. s^I_{tT} represents the additional risk incorporated in the index rate. For the index rates we have

$$\begin{aligned} I(T;t) &= \frac{1}{\alpha_{I,T}}\left(\exp\left(\int_T^{m_I(T)} f^I(t,\xi)d\xi\right) - 1\right) \\ &= \frac{1}{\alpha_{I,T}}\left(\exp\left(\int_T^{m_I(T)} (f(t,\xi) + s^I(t,\xi))d\xi\right) - 1\right) \end{aligned} \tag{92}$$

We can also define the discretely compounded spread of the index rate over funding as

$$S^I(t,T) = I(t,T) - \frac{1}{\alpha_{I,T}}\left(\frac{P(t,T)}{P(t,m_I(T))} - 1\right).$$

Aside from the fixing of the reference index on the day, which gives the relative discount factor from its spot date to its maturity date, P^I is not directly observable in the markets, as standard instruments are subject to collateralisation, whereas the index curve assumes an uncollateralised investment. Furthermore, even if we had uncollateralised market instruments available, these would be with a fixed counterparty, whereas the reference index represents the average of a panel of institutions. Not only would the counterparty not be an exact match, but the panel composition typically changes over time to keep the credit quality at a given, high level, which makes the implied credit risk of the reference index lower than that of any particular market counterparty over longer time periods.

13.2 Foreign Currency Analogy

We therefore need to consider the pricing of collateralised instruments referencing the index, for which we need to define *processes* for the index rates, or equivalently for rolling investments in the index. To do this, we use a foreign currency analogy. We consider a unit of currency borrowed on a rolling basis at the standard discount rate and invested in the index rate from an arbitrary but fixed time t_0 as a "currency" for investments into I, the price of which at time t we denote by X^I_t. This is a formal quantity, which captures the credit history of an investment in I up to time t; it will allow us to treat the P^I as discount bond prices at any point in time without needing to consider the credit history[5]. The only constraint we need

[5]Alternatively, we could have considered an abstract counterparty for investments at I and discount bonds "issued" by it, with the additional assumption that it is possible to separate out the credit risk as a common factor. The cross-currency analogy appeals more to us, as we can simply think of an obligation linked to I as a separate, but closely linked currency.

to impose is that $X_t^I > 0$ almost surely for any t, which is not an unreasonable assumption, since a total loss of a rolling deposit investment without any recovery is extremely unlikely. We will assume without loss of generality that $t_0 = 0$ and therefore $X_0^I = 1$.

As X_t^I models the cumulative credit risk of an investment at the index rate I, it must be a non-increasing process, and it should clearly move by jumps only. The measure change between the risk-neutral measures for the base currency and the index "currency" is given by

$$\tilde{X}_t^I = X_t^I \exp\left(\int_0^t (f_{\xi\xi}^I - f_{\xi\xi})d\xi\right) = X_t^I \exp\left(\int_0^t s_{\xi\xi}^I d\xi\right);$$

in financial terms $\tilde{X}_t^I$ is the value at time t of a rolling investment at I funded by rolling borrowing at the funding rate beginning with one unit of currency at time 0. $\tilde{X}_t^I$ must be a martingale in any reasonable model we build. s_{tt}^I is its instantaneous growth in the absence of defaults, which is therefore the compensator of the jumps of X_t^I under the risk-neutral measure. With the additional assumption of a recovery rate (or jump size) process we could imply a default density (or hazard rate) from s_{tt}^I; this is a concrete modelling decision which we want to avoid for the time being. We simply assume that $\tilde{X}_t^I$ is a martingale in the risk-neutral measure; as it is clearly non-decreasing on the continuous parts of its paths, it must be quadratic pure jump.

The foreign currency analogy allows us to absorb credit events in the exchange rate X^I. We can therefore treat the index discount curve P^I as continuous and continuously evolving in line with standard interest rate modelling practice; dynamics with jumps are possible in this framework, but we will not consider this here. We therefore make the additional assumption that $P^I(t, T)$ is a continuous stochastic process for each T. Therefore the HJM framework should apply, and we can assume that both the discount curve and the index curve satisfy the HJM assumptions.

Because the jumps are concentrated in the exchange rate X^I and all other components are pure diffusion, we can expect that jumps do not contribute to the pricing of market instruments and that we hence only need to consider the compensators, i.e. the spreads. The remainder of this section is devoted to proving this and to deriving explicit pricing formulae for vanilla instruments.

13.3 Pricing Collateralised FRAs

We now consider a forward rate agreement in a collateralised setting where we receive the index rate minus a fixed rate K from T to $U = m_I(T)$. We assume that the agreement is physically settled and fully collateralised, i.e. at time T the trade starts accruing $I(T;T) - K$ until final settlement at U with continuing full collateralisation. The value at T is therefore[6]

$$\mathrm{FRA}_c{}^I(K,T;T) = \alpha_{I,T}(I(T,T) - K)P(t,U) \tag{93}$$

In our cross-currency analogy this is a quanto payment of the index "currency" interest in our base currency.

Under the terminal measure for the index "currency" associated to the numeraire $P^I(\cdot,U)$, which we will denote by $\mathbb{P}^I_U$, the index rate $I(\cdot,T)$ is a martingale. The Girsanov-Meyer Theorem (see e.g. [*Protter*, 2004, Ch. III, Th. 35]) gives the corresponding dynamics under the terminal measure for the actual currency associated to the numeraire $P(\cdot,U)$, which we denote by $\mathbb{P}_U$. Denoting the forward exchange rate for a time τ by

$$X_t^{I,\tau} = X_t^I \frac{P^I(t,\tau)}{P(t,\tau)} = X_t^I \exp\left(-\int_t^\tau s^I(t,\xi)d\xi\right), \tag{94}$$

the measure change is given by $Z_t = \frac{X_t^{I,U}}{X_0^{I,U}}$ and we obtain from the Girsanov-Meyer Theorem that

$$\mathbb{E}_U\left[I(T,T)|\mathcal{F}_t\right] = I(t,T) + \mathbb{E}_U\left[\int_t^T Z_s d\Big[Z^{-1}, I(\cdot,T)\Big]_s \middle| \mathcal{F}_t\right]$$

and therefore for the value of the FRA

$$\mathrm{FRA}_c{}^I(K,T;t) = \alpha_{I,T}P(t,U)(I(t,T) - K + C^I(t,T)) \tag{95}$$

where

$$C^I(T;t) = \mathbb{E}_U\left[\int_t^T Z_s d\Big[Z^{-1}, I(\cdot,T)\Big]_s \middle| \mathcal{F}_t\right] \tag{96}$$

is the "quanto" convexity correction. This shows that the collateralised FRA can be priced by discounting the intrinsic value with the collateralised discount rate and adding a convexity correction factor. Intuitively, the change of discounting is the "first order" correction of the price, while the convexity should be a "second order" correction, determined by the covariance of the forward "exchange" rate with the forward index rate. We will now analyse the convexity correction $C^I(t,T)$ further to substantiate this intuition.

By the integration by parts formula

$$\begin{aligned} Z_t^{-1} &= \frac{P^I(0,U)}{P(0,U)}(X_t^I)^{-1}\frac{P(t,U)}{P^I(t,U)} \\ &= \frac{P^I(0,U)}{P(0,U)}\left(\int_0^t (X_-^I)^{-1} d\frac{P(\cdot,U)}{P^I(\cdot,U)} + \int_0^t \frac{P(\cdot,U)}{P^I(\cdot,U)} d(X^I)^{-1} + \left[(X^I)^{-1}, \frac{P(\cdot,U)}{P^I(\cdot,U)}\right]_t\right). \end{aligned}$$

[6]More precisely the value is already determined on the fixing date of $I(T)$, but we prefer to keep notation simple and therefore do not distinguish fixing and value dates here.

The quadratic covariance term is constant because X^I is quadratic pure jump while $\frac{P(\cdot,U)}{P^I(\cdot,U)}$ is continuous, hence $\left[\left[(X^I)^{-1}, \frac{P(\cdot,U)}{P^I(\cdot,U)}\right], I(\cdot,T)\right]$ is constant. Similarly, $\left[(X^I)^{-1}, I(\cdot,T)\right]$ is constant, so

$$\left[\int_0^t \frac{P(\cdot,U)}{P^I(\cdot,U)} d(X^I)^{-1}, I(\cdot,T)\right] = \int_0^t \frac{P(\cdot,U)}{P^I(\cdot,U)} d[(X^I)^{-1}, I(\cdot,T)] = 0.$$

We therefore have

$$\begin{aligned}
\int_t^T Z_s d[Z^{-1}, I(\cdot,T)]_s &= \int_t^T X_s^I \frac{P^I(s,U)}{P(s,U)} d\left[\int_0^\cdot (X_-^I)^{-1} d\frac{P(\cdot,U)}{P^I(\cdot,U)}, I(\cdot,T)\right]_s \\
&= \int_t^T \frac{X_s^I}{X_{s-}^I} \frac{P^I(s,U)}{P(s,U)} d\left[\frac{P(\cdot,U)}{P^I(\cdot,U)}, I(\cdot,T)\right]_s \\
&= \int_t^T \frac{P^I(s,U)}{P(s,U)} d\left[\frac{P(\cdot,U)}{P^I(\cdot,U)}, I(\cdot,T)\right]_s \qquad (97)
\end{aligned}$$

where in the last step we used the fact that the integrator is continuous.

We can now make use of our basic assumption that both the discount and the index curve evolve continuously and hence satisfy the HJM assumptions. Under the forward measure $\mathbb{P}_U$ we have

$$\frac{P(t,T)}{P(t,U)} = \frac{P(0,T)}{P(0,U)} \exp\left(\int_0^t (\Sigma_{uT} - \Sigma_{uU}) dW_u - \frac{1}{2}\int_0^t \|\Sigma_{uT} - \Sigma_{uU}\|^2 du\right)$$

where $\Sigma(\tau;t) \equiv \Sigma_{t\tau}$ is the HJM volatility process of the discount bond[7] expiring at τ and W is a (d-dimensional) Brownian motion under $\mathbb{P}_U$. Also, we can assume without loss of generality that

$$\frac{P^I(t,T)}{P(t,Ut)} = \frac{P^I(0,T)}{P(0,U)} \exp\left(\int_0^t (\Sigma_{uT}^I - \Sigma_{uU}) dW_u + \int_0^t \beta_u^T du\right)$$

for some drift process β^T. By the Girsanov-Meyer Theorem and arguments analogous to the derivation of (97) for the pure jump component

$$\begin{aligned}
W_t^I = W_t - \int_0^t Z_u^{-1} d[Z,W] &= W_t - \int_0^t \frac{P(u,U)}{P^I(u,U)} d\left[\frac{P^I(\cdot,U)}{P(\cdot,U)}, W\right]_u \\
= W_t - d\left[\int_0^t (\Sigma_{uU}^I - \Sigma_{uU}) dW_u, W\right]_u &= W_t + \int_0^t (\Sigma_{uU} - \Sigma_{uU}^I)^* du
\end{aligned}$$

[7]Recall that in HJM $\Sigma_{t\tau} = -\int_t^\tau \sigma_{tu} du$ where σ_{tu} is the volatility of the instantaneous forward rate f_{tu}.

is a Brownian motion under $\mathbb{P}^I_U$. As $\frac{P^I(\cdot,T)}{P^I(\cdot,U)}$ is a tradeable under that measure it must be an exponential martingale, so

$$\frac{P^I(t,T)}{P^I(t,U)} = \frac{P^I(0,T)}{P^I(0,U)} \exp\left(\int_0^t (\Sigma^I_{uT} - \Sigma^I_{uU}) dW^I_u - \frac{1}{2}\int_0^t \left\|\Sigma^I_{uT} - \Sigma^I_{uU}\right\|^2 du\right).$$

Re-writing in terms of W we obtain

$$\begin{aligned}\frac{P^I(t,T)}{P^I(t,U)} =& \frac{P^I(0,T)}{P^I(0,U)} \exp\left(\int_0^t (\Sigma^I_{uT} - \Sigma^I_{uU}) dW_u - \frac{1}{2}\int_0^t \left\|\Sigma^I_{uT} - \Sigma^I_{uU}\right\|^2 du\right) \\ & \cdot \exp\left(\int_0^t (\Sigma^I_{uT} - \Sigma^I_{uU})(\Sigma_{uU} - \Sigma^I_{uU})^* du\right)\end{aligned} \tag{98}$$

Inserting these definitions and the definition of $I(\cdot,T)$ in (91) into (97) we obtain

$$\int_t^T Z_s d\Big[Z^{-1}, I(T)\Big]_s = \frac{1}{\alpha_{I,T}} \int_t^T \frac{P^I(T;u)}{P^I(U;u)} (\Sigma^I_{uT} - \Sigma^I_{uU})(\Sigma_{uU} - \Sigma^I_{uU})^* du.$$

In general, the $\Sigma_{t\tau}$ and $\Sigma^I_{t\tau}$ will be stochastic, and an explicit computation of the convexity adjustment will not be possible. However, if the model is Gaussian, i.e. $\Sigma_{t\tau}$ and $\Sigma^I_{t\tau}$ are deterministic, then we can use Fubini and (98) to compute

$$\begin{aligned} C^I(t,T) &= \frac{1}{\alpha_{I,T}} \int_t^T \mathbb{E}_U\left[\left.\frac{P^I(u,T)}{P^I(u,U)}\right| \mathcal{F}_t\right] (\Sigma^I_{uT} - \Sigma^I_{uU})(\Sigma_{uU} - \Sigma^I_{uU})^* du \\ &= \frac{1}{\alpha_{I,T}} \frac{P^I(t,T)}{P^I(t,U)} \int_t^T \exp\left(\int_t^u (\Sigma^I_{\xi T} - \Sigma^I_{\xi U})(\Sigma_{\xi U} - \Sigma^I_{\xi U})^* d\xi\right) (\Sigma^I_{uT} - \Sigma^I_{uU})(\Sigma_{uU} - \Sigma^I_{uU})^* du \\ &= \left(I(t,T) + \frac{1}{\alpha_{I,T}}\right)\left(\exp\left(\int_t^T (\Sigma^I_{uT} - \Sigma^I_{uU})(\Sigma_{uU} - \Sigma^I_{uU})^* du\right) - 1\right) \end{aligned}$$

This is as expected; the forward index rate $I(T)$ is shifted log-normal in a Gaussian HJM model with shift $\frac{1}{\alpha_{I,T}}$ and, separating out the jump martingale component, the forward "exchange" rate is log-normal as well, and the exponential above is simple the normalised covariance between the two log-normal variables. $\left\|\Sigma^I_{tT} - \Sigma^I_{tU}\right\|$ is the instantaneous volatility of $I(t,T)$ under the shifted law, the equivalent log-normal volatility is approximately $\frac{I(t,T)}{I(t,T)+\frac{1}{\alpha_{I,T}}}$ times the shifted log-normal volatility. Hence if the covariance is not too large, so that $\exp(x) \approx 1 + x$ is a reasonable approximation, we have

$$C^I(t,T) \approx I(t,T)\rho\bar{\sigma}_X\bar{\sigma}_{I,T}(T-t) \tag{99}$$

where $\bar{\sigma}_{I,T}$ is the implied volatility of $I(T)$, $\bar{\sigma}_X$ the implied volatility of the forward "exchange" rate at U and ρ the (term) correlation between them.

While formula (99) is approximate and depends on the assumption of a Gaussian HJM model, it can be used as an estimate of the convexity adjustment. Since the forward "exchange" rate (excluding the jump component) is the exponential of the cumulative spread, its volatility and the correlation with the index rate can be implied from the behaviour of the spread.

13.4 Pricing Collateralised Swaps

The discussion of collateralised FRA pricing generalises directly to collateralised swap pricing. Consider a swap struck at K linked to a reference index rate I; we assume the fixed leg accrues between dates $T_0, \ldots, T_n$ with accrual fractions α_i and the floating leg accrues between the dates $T_0 = t_0, \ldots, t_m = T_n$ on the reference index rate with value date at the beginning of each period with accrual factors $\hat{\alpha}_j$. The valuation of the fixed leg simply discounts the cash flows on the discount curve, i.e. defining the swap annuity as

$$B_t = \sum_{i=1}^{n} \alpha_i P(t, T_i)$$

the value of the fixed leg at time t is KB_t.

The floating leg works basically (up to small timing differences) as a collection of collateralised forward rate agreements; the jth flow can be valued in the t_j-forward measure and has value

$$\hat{\alpha}_j P(t, t_j)\mathbb{E}_{t_j}\left[I(t_{j-1}, t_{j-1}) \middle| \mathcal{F}_t \right] = \hat{\alpha}_j P(t, t_j)(I(t, t_{j-1}) + C^I(t, t_{j-1}))$$

so that the floating leg value is

$$\sum_{j=1}^{m} \hat{\alpha}_j P(t, t_j)(I(t, t_{j-1}) + C^I(t, t_{j-1}))$$

Neglecting small timing differences between the accrual periods of the swap and the reference index rate, we can also write this as

$$P(t, t_0) - P(t, t_n) + \sum_{j=1}^{m} \hat{\alpha}_j P(t, t_j)(S^I(t, t_{j-1}) + C^I(t, t_{j-1}))$$

which represents the value as a standard floating leg on the discount curve plus the additional value from the spread and the convexity. As discussed above, it should be reasonable to assume that the convexity corrections C^I are zero in normal market conditions, but in stressed market situations they are likely to become material.

13.5 Interpreting Market Quotes and General Pricing

To correctly interpret market quotes, it is first necessary to construct the market standard discount curve. For US Dollars the reference overnight rate is the Fed Funds Effective Rate, as it is normally used as the reference rate in collateral agreements[8]; a corresponding discount curve can be constructed from the Overnight Index Swap market up to maturities of several years, and extended with some basic assumptions.

In other currencies the standard discount curve needs to be established; this depends on the conventions of standard collateralisation agreements and the discount rates referenced therein. If an overnight rate in the same currency is used, such as EONIA for Euros, then this discount curve needs to be constructed in the same manner. If the swaps market in the currency is dominantly collateralised using a different currency, such as US Dollars (or Euros), then the correct discount curve to use is the cross-currency funding curve constructed from the funding currency discount curve and FX forwards and swaps. It may be that different standard discount curves are necessary for different types of collateral agreements, and that market quotes need to be selected to have a consistent underlying discount rate assumption.

Once the standard market discount curve has been constructed, the index curves for all reference indices traded in the market need to be constructed from standard market quotes (discount rate fixings, FRAs, swaps, basis swaps) using the valuation formulae above. The collection of all these curves then needs to be used in calculating the market price of collateralised trades; note that, as discussed above, this is the unwind value of trades and does not represent the actual value to the bank because a market standard marginal funding cost is assumed.

Pricing of more general, non-vanilla positions needs to use the same principles as above and adapt the relevant pricing model to correctly account for funding. Where assets are held and can be used in repo transactions for cheaper funding, this must be taken into account, as for example in equity derivative pricing (see *Piterbarg* [2010]). The treatment is pricing model specific and needs to be addressed model by model.

A The Markov Property and Separable Volatility Functions

This section provides a proof that for a Gaussian HJM-model to admit a Markovian short rate process, the volatility function must be separable, i.e. a product of a

[8] The London Clearing House recognised the Fed Funds Effective Rate as the standard discount rate by switching to OIS discounting for margin calls in mid-2010.

purely time-dependent and a purely expiry-dependent function. The proof applies to more general processes than (33) of the form

$$R_t = R_0 + \int_0^t \alpha(u,t)du + \int_0^t \sigma(u,t)dW_u \tag{100}$$

where $R_0 \in \mathbb{R}^k$, $\alpha : C \to \mathbb{R}^k$ and $\sigma : C \to L(\mathbb{R}^d, \mathbb{R}^k)$ are (measurable) functions, $C = \{(u,t)|0 \leq u \leq t \leq T^*\}$, and W is a d-dimensional Brownian motion. The proof shows, for example, that a multi-factor generalised Vasicek model must have a separable volatility term structure as a Gaussian HJM model. The proof follows the idea of [*Musiela and Rutkowski*, 1997, Prop. 13.3.2] but provides some corrections, not least to the statement of the result.

Proposition A.1. *Let R be a k-dimensional random process with law (100) and $k \leq d$. If σ is non-degenerate, i.e. $\sigma(t,T)$ has maximal rank almost everywhere, then R is Markovian if, and only if, there exist functions $g : [0,T^*] \to L(\mathbb{R}^d, \mathbb{R}^k)$ and $h : [0,T^*] \to L(\mathbb{R}^k, \mathbb{R}^k)$ such that $\sigma(t,T) = h(T)g(t)$. Such a volatility function σ is called* separable.

Proof. As α and σ are deterministic, R_t is Markovian if, and only if,

$$X_t = \int_0^t \sigma(u,t)dW_u$$

is Markovian, which is equivalent to

$$\mathbb{E}\left[f(X_s)|\mathcal{F}_t\right] = \mathbb{E}\left[f(X_s)|X_t\right] \qquad \forall t \leq s \leq T^* \tag{101}$$

for any bounded Borel-measurable $f : \mathbb{R}^k \to \mathbb{R}$. As

$$\begin{aligned} X_s &= X_t + \int_0^s \sigma(u,s)dW_u - \int_0^t \sigma(u,t)dW_u \\ &= X_t + \int_t^s \sigma(u,s)dW_u + \int_0^t (\sigma(u,s) - \sigma(u,t))dW_u \end{aligned}$$

and $\int_t^s \sigma(u,s)dW_u$ is independent of $\mathcal{F}_t$, (101) holds if, and only if, given the random variable X_t

$$I(t,s) = \int_0^t \left(\sigma(u,s) - \sigma(u,t)\right) dW_u$$

depends only on the increments of W on $[t,s]$. But $I(t,s)$ is, by definition of the Itô integral, independent of those increments.

Hence if R_t is Markovian, then $I(t,s)$ is completely determined by X_t. As the joint law of $(I(t,s), X_t)$ is Gaussian centred at 0, this implies that there exists a linear map $A(t,s) \in L(\mathbb{R}^k, \mathbb{R}^k)$ such that $I(t,s) = A(t,s)X_t$. Consequently

$$\int_0^t (\sigma(u,s) - \sigma(u,t))\, dW_u = A(t,s) \int_0^t \sigma(u,t) dW_u = \int_0^t A(t,s)\sigma(u,t) dW_u.$$

As σ is non-degenerate, it follows that $\sigma(u,s) = (A(t,s) + I)\,\sigma(u,t)$ for all $0 \le u \le t \le s \le T^*$. As σ has maximal rank κ everywhere and $\kappa \le d$, $I + A(t,s)$ is invertible for all $0 \le t \le s \le T^*$. Setting $h(T) = (A(T,T^*) + I)^{-1}$ and $g(t) = \sigma(t,T^*)$, it follows that

$$\sigma(t,T) = (A(T,T^*) + I)^{-1}\, \sigma(t,T^*) = h(T)g(t).$$

Conversely, if σ is separable, then $X_t = h(t) \int_0^t g(u) dW_u$.

$$I(t,s) = (h(s) - h(t)) \int_0^t g(u) dW_u = (h(s)h(t)^{-1} - I)X_t$$

is completely determined by X_t, and hence independent of the increments of W on $[t,s]$. Therefore X is Markovian. □

It is often desirable to have term structures that are time homogeneous, i.e. they only depend on the time to expiry. For a separable volatility function $\sigma(t,T) = h(T)g(t)$ to be time homogeneous, there must exist a function $\tilde{\sigma} : [0,T^*] \to L(\mathbb{R}^k, \mathbb{R}^d)$ such that $\sigma(t,T) = \tilde{\sigma}(T-t)$. If g and h are differentiable, this implies $h'(T)g(t) + h(T)g'(t) = 0$ for all $0 \le t \le T$, so if σ is non-degenerate and $k \le d$ it follows that $h^{-1}(T)h'(T) = \text{const}$, hence $h(T) = e^{AT}$ for some $A \in L(\mathbb{R}^k, \mathbb{R}^k)$. Similarly, if σ is non-degenerate and $k \ge d$, then $g(t) = e^{-Bt}$, and for $d = k$ we must have $B = A$ and $\tilde{\sigma} = e^{A(T-t)}$.

References

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

Asmussen, S., and P. W. Glynn, *Stochastic Simulation: Algorithms and Analysis*, Springer-Verlag, New York, 2007.

Baxter, M. W., General interest rate models and the universality of HJM, in *Mathematics of Derivative Securities*, edited by S. Pliska and M. A. H. Dempster, Cambridge University Press, Cambridge, 1997.

Belomestny, D., C. Bender, and J. Schoenmakers, True upper bounds for bermudan products via non-nested monte carlo, *Mathematical Finance*, *19*, 53–71, 2009.

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

Heath, D. C., R. A. Jarrow, and A. Morton, Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation, *Econometrica*, *60*, 77–105, 1992.

Hunt, P., J. Kennedy, and A. Pelsser, Markov-functional interest rate models, *Finance and Stochastics*, *4*, 391–408, 2000.

Hunt, P. J., and J. E. Kennedy, *Financial Derivatives in Theory and Practice*, Wiley Series in Probability and Statistics, revised ed., Wiley, Chichester, 2004.

Hunter, C. S., P. Jckel, and M. S. Joshi, Getting the drift, *Risk*, *14(July)*, 81–84, 2001.

Jäckel, P., Using a non-recombining tree to design a new pricing method for bermudan swaptions, *QuaRC working paper, The Royal Bank of Scotland*, 2000.

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

Joshi, M. S., and R. Tang, Effective sub-simulation-free upper bounds for the monte carlo pricing of callable derivatives and various improvements to existing methodologies, *SSRN Preprint*, (2095988), 2012.

Kloeden, P. E., and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992.

Longstaff, F., and E. S. Schwartz, Valuing american options by simulation: a least-squares approach, *Review of Financial Studies*, *14*, 113–147, 2001.

Musiela, M., and M. Rutkowski, *Martingale Methods in Financial Modelling*, Springer-Verlag, Berlin, 1997.

Piterbarg, V., Funding beyond discounting: collateral agreements and derivatives pricing, *Risk*, *23*(2), 2010.

Protter, P., *Stochastic Integration and Differential Equations*, 2nd ed., Springer-Verlag, Berlin, Heidelberg, New York, 2004.

Rogers, L. C. G., Monte carlo valuation of american options, *Mathematical Finance*, *12*, 271–286, 2002.

Toepler, M., Markov functional interest rate models with stochastic volatility, Master's thesis, Oxford University, 2009.