Interest rate modeling
Market models, products and risk management
(following [AP10-1], [AP10-2] and [AP10-3])

Alan Marc Watson

July 5, 2016

Abstract

This document contains a brief summary of Andersen and Piterbarg's superb three-volume treatise on fixed-income derivatives. I have used this as a self-study guide and also to document my subsequent model implementations

Contents

1 Fundamentals of interest rate modeling **6**
1.1 Fixed income notations 6
1.2 Fixed income probability measures 6
1.2.1 Risk neutral measure 6
1.2.2 T-forward measure 7
1.2.3 Spot measure 7
1.2.4 Swap measures 7
1.3 The HJM analysis 8
1.3.1 The Gaussian model 9
1.3.2 Gaussian model with Markovian short rate 10

2 Vanilla models with local volatility **11**
2.1 General framework 11
2.2 The CEV model 11
2.2.1 Valuation of European call options in the CEV model 12
2.2.2 Regularization 12
2.2.3 Displaced diffusion models 12
2.3 Finite difference solutions 13
2.3.1 Multiple λ and T 13
2.4 Forward equation for call options 13
2.5 Time-dependent local volatility function φ 14
2.5.1 Separable case 14
2.5.2 Skew averaging 14

3 Vanilla models with stochastic volatility **17**
3.1 Basic properties 17
3.2 Fourier integration 17
3.2.1 Calculation of option prices when MGF is available for $\ln(S(t))$. . . . 17
3.2.2 Direct integration 18

3.2.3 Fourier integration for arbitrary European Payoffs . . . 19
3.3 Averaging methods for models with time-dependent parameters . . . 21
3.3.1 Volatility averaging . . . 21
3.3.2 Skew averaging . . . 22
3.3.3 Volatility of variance averaging . . . 23
3.3.4 Calibration by parameter averaging . . . 23

4 One-factor short rate models **26**
4.1 Review of Swaps and swaptions . . . 26
4.2 The Ho-Lee model . . . 27
4.3 The mean-reverting GSR model . . . 27
4.3.1 Swaption pricing . . . 29
4.3.2 Swaption calibration . . . 31
4.3.3 Monte Carlo simulation . . . 32
4.4 The affine one-factor model . . . 33
4.4.1 Bond pricing . . . 34
4.4.2 Discount bond calibration . . . 34
4.4.3 European option pricing . . . 35
4.4.4 Swaption calibration . . . 37
4.5 Log-normal short-rate models . . . 37
4.6 Spanned and unspanned stochastic volatility . . . 38

5 Multi-factor short-rate models **39**
5.1 The Gaussian multi-factor model . . . 39
5.2 Two-factor Gaussian model . . . 41
5.3 Swaption pricing (via Gaussian swap rate approximation) . . . 41
5.3.1 Calibration . . . 42
5.3.2 Monte Carlo simulation . . . 43

6 The one-factor quasi-Gaussian model **45**
6.1 Definition . . . 45
6.2 Local volatility . . . 46
6.2.1 Linear local volatility . . . 48
6.2.2 Volatility calibration . . . 50
6.2.3 Computational tip . . . 51
6.2.4 Mean reversion calibration . . . 51
6.3 Stochastic volatility . . . 51
6.4 Quasi-Gaussian multi-factor models . . . 52

7 The Libor market model **53**
7.1 The basics . . . 53
7.1.1 Probability measures . . . 53
7.1.2 Link to HJM . . . 54
7.1.3 Choice of $\sigma_n(t)$ . . . 54
7.2 Correlation structure . . . 55
7.2.1 Empirical PCA . . . 55
7.2.2 Empirical estimates for forward rate correlations . . . 56
7.3 Pricing European options . . . 56
7.3.1 Caplets . . . 56
7.3.2 Swaptions . . . 57
7.3.3 Spread options . . . 60
7.4 Calibration . . . 61
7.4.1 Calibration objective function . . . 61

7.4.2 Calibration algorithm . . . 62
7.4.3 Correlation calibration to spread options . . . 64
7.5 Monte Carlo simulation . . . 65
7.5.1 Euler-type schemes . . . 65
7.5.2 Special-purpose schemes with drift Predictor-Corrector . . . 65
7.6 Interpolation of front / back stubs . . . 66
7.6.1 Back stub interpolation . . . 66
7.6.2 Back stub interpolation inspired by the Gaussian model . . . 67
7.6.3 Front stub interpolation inspired by the Gaussian model . . . 68
7.7 Advanced approximation for swaption pricing via Markovian projection . . . 68
7.7.1 Advanced formula for the swap rate volatility . . . 70
7.7.2 Advanced formula for the swap rate skew . . . 70
7.7.3 Skew and smile calibration . . . 70
7.8 Other extensions . . . 72
7.8.1 Evolving separate discount and forward rate curves . . . 72
7.8.2 SV models with non-zero correlation . . . 72
7.8.3 Multi-stochastic volatility extensions . . . 72

8 Single-rate vanilla derivatives **73**
8.1 European swaptions . . . 73
8.2 Caps and floors . . . 74
8.3 Terminal swap rate (TSR) models . . . 74
8.3.1 Linear TSR models . . . 75
8.4 Libor-in-Arrears . . . 76
8.5 Libor-with-Delay . . . 76
8.6 CMS and CMS-Linked cash flows . . . 77
8.6.1 Linear TSR model . . . 78
8.6.2 Libor market model . . . 78
8.6.3 Quasi-Gaussian model . . . 79
8.6.4 Correcting non-arbitrage-free methods . . . 79
8.6.5 CDF and PDF of CMS rate in Forward measure . . . 80
8.6.6 SV model for CMS rate . . . 81
8.6.7 Dynamics of CMS rate in forward measure . . . 82
8.7 Quanto CMS . . . 83
8.8 Eurodollar futures . . . 84
8.8.1 The problem and the plan . . . 85
8.8.2 Expansion around futures value . . . 85
8.8.3 Forward rate variances . . . 86
8.8.4 Forward rate correlations . . . 86

9 Multi-rate vanilla derivatives **87**
9.1 Dependence structure via copulas . . . 87
9.2 Gaussian copula method . . . 87
9.3 General couplas . . . 88
9.4 Copula methods for spread options . . . 89
9.4.1 Power gaussian . . . 90
9.5 Limitations of the copula method . . . 90

10 Callable Libor Exotics **92**
10.1 Review of basic notions . . . 92
10.2 Model calibration for CLE's . . . 93
10.3 Valuation theory . . . 93
10.4 Monte Carlo valuation . . . 94
10.4.1 Regression for the underlying . . . 95
10.4.2 Using regressed variables for decision only . . . 96
10.4.3 Controlling the MC estimate bias . . . 97
10.4.4 Upper bound . . . 97
10.4.5 Choice of regression variables . . . 97
10.5 Implementation . . . 98
10.6 Valuation with low-dimensional models . . . 99
10.7 Suitable analogue for core swap rates . . . 100

11 Bermudan swaptions **102**
11.1 Local projection method . . . 102
11.2 Amortizing, accreting and other non-standard Bermudan swaptions . . . 103
11.2.1 Relating non-standard to standard swap rates . . . 103
11.2.2 Representative swaption approach (payoff matching) . . . 104
11.2.3 Basket approach . . . 105
11.2.4 American swaptions . . . 106
11.3 Flexi-swaps . . . 106
11.3.1 Purely local bounds . . . 107
11.4 Fast pricing via exercise premia representation . . . 107

12 TARN's, volatility swaps and other derivatives **108**
12.1 TARN's (targeted redemption notes) . . . 108
12.1.1 Local projection method . . . 108
12.1.2 Effect of volatility smiles . . . 109
12.2 Volatility swaps . . . 109
12.2.1 Volatility swaps with shout option . . . 109
12.2.2 Min-Max volatility swaps . . . 110
12.3 Forward swaption straddles . . . 110

13 Risk management **111**
13.1 Value at risk . . . 111

14 Payoff smoothing for Monte Carlo **113**
14.1 Tube Monte Carlo for digital options . . . 113
14.2 Tube Monte Carlo for barrier options . . . 114

15 Pathwise differentiation **116**
15.1 On the unsuitability of the likelihood ratio method in interest rate modeling . 117
15.2 CLE's . . . 118
15.3 Barrier options . . . 118
15.3.1 Digital option warm-up . . . 119
15.3.2 Barrier options . . . 119
15.4 Pathwise differentiation for Monte Carlo based models . . . 120

16 Importance sampling and control variates **121**
16.1 Importance sampling 121
16.2 Payoff smoothing via importance sampling 124
16.2.1 Binary options 124
16.2.2 TARN's 124
16.3 Control variates 126
16.3.1 Model-based control variates 128
16.3.2 Instrument-based control variates 130
16.3.3 Dynamic control variates 131

1 Fundamentals of interest rate modeling

1.1 Fixed income notations

Some notations:

- $P(t,T)$: time-t price of a zero-coupon bond (ZCB) delivering \$1 at time $T \geq t$.
- $P(t,T,T+\tau) = \frac{P(t,T+\tau)}{P(t,T)}$: time-t forward price for the ZCB spanning $[T,T+\tau]$[1].
- $y(t,T,T+\tau)$: continuously compounded yield, defined by

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

- $L(t,T,T+\tau)$ simple forward rate, defined by

$$1+\tau L(t,T,T+\tau) = \frac{1}{P(t,T,T+\tau)} \Longrightarrow L(t,T,T+\tau) = \frac{1}{\tau}\left[\frac{P(t,T)}{P(t,T+\tau)} - 1\right]$$

  $L(t,T,T+\tau)$ will generally be the *Libor* rates quoted in the interbank market.
- $f(t,T)$ is the time-t instantaneous forward rate

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

- Relation between $f(t,T)$ and $P(t,T,T+\tau)$

$$\begin{aligned} P(t,T,T+\tau) &= \exp\left(-\int_T^{T+\tau} f(t,u)\,du\right), \qquad f(t,T) = -\frac{\partial}{\partial T}\ln P(t,T) \\ y(t,T,T+\tau) &= \frac{1}{\tau}\int_T^{T+\tau} f(t,u)\,du, \qquad f(t,T) = \lim_{\tau\to 0} y(t,T,T+\tau) \end{aligned}$$

- $r(t) = f(t,t)$ short/spot rate.
- $F(t,T,T+\tau)$ futures rate.

1.2 Fixed income probability measures

We outline some of the main measures used in the study of fixed-income and list some of their properties. Let $V(t)$ denote the time-t value of a derivative security with payout function $V(T)$.

1.2.1 Risk neutral measure

The numraire defining the measure Q is the continuously compounded money market account $\beta(t)$.

$$\begin{cases} d\beta(t) = \beta(t)dt \\ \beta(0) = 1 \end{cases} \Longrightarrow \beta(t) = e^{\int_0^t r(u)\,du}$$

In the absence of arbitrage the numraire deflated process $\frac{V(t)}{\beta(t)}$ must be a martingale, which yields the risk-neutral pricing formula

$$V(t) = E_t^Q\left[V(T)e^{-\int_t^T r(u)\,du}\right]$$

[1] This is the time-T purchase price of a bond maturing at $T+\tau$, fixed at time $t<T$. To see this, at time t we can buy a $(T+t)$-maturity bond for $P(t,T+\tau)$ and short $\frac{P(t,T+\tau)}{P(t,T)}$ T-maturity bonds, which costs us 0. At time T, we close the short position on the T-maturity bond by paying $\frac{P(t,T+\tau)}{P(t,T)}\cdot P(T,T)$ to the buyer, and at time $T+\tau$ we receive \$1 for our $(T+t)$-maturity bond.

In particular, for $V(T) = 1$ we obtain the price of a ZCB

$$P(t,T) = E_t^Q\left[e^{-\int_t^T r(u)\,du}\right]$$

1.2.2 T-forward measure

The T-forward measure Q^T uses a T-maturity ZCB as numraire. The risk-neutral pricing formula in this case reads

$$V(t) = P(t,T)E_t^T\left[V(T)\right]$$

In the absence of arbitrage, the forward Libor rate $L(t,T,T+\tau)$ is a Q^T-martingale such that

$$L(t,T,T+\tau) = E_t^{T+\tau}\left[L(T,T,T+\tau)\right], \quad t \leq T$$

1.2.3 Spot measure

Consider a tenor structure $0 < T_0 < \cdots < T_N$. Sometimes it is convenient to introduce a numraire that can be extended to arbitrary horizons by compounding. Consider the following situation:

(i) Time 0: invest \$1 in $\frac{1}{P(0,T_1)}$ T_1-maturity discount bonds, returning an amount

$$\frac{1}{P(0,T_1)} = 1 + \tau_0 L(0,0,T_1)$$

(ii) Time T_1: Re-invest this amount in T_2-maturity bonds, returning

$$\frac{1}{P(0,T_1)}\frac{1}{P(T_1,T_2)} = [1 + \tau_0 L(0,0,T_1)][1 + \tau_1 L(T_1,T_1,T_2)]$$

(iii) Iterating this we obtain a process

$$B(t) = P(t,T_{i+1})\prod_{n=0}^{i}\left[1 + \tau_n L_n(T_n)\right], \qquad T_i < t \leq T_{i+1}$$

The **spot measure** Q^B is the measure induced by this process and the time-t value of a secutiry in this measure is by definition

$$V(t) = E_t^B\left[V(T)\frac{B(t)}{B(T)}\right], \quad \frac{B(t)}{B(T)} = \prod_{n=i+1}^{j}\left[1 + \tau_n L_n(T_n)\right]^{-1}\frac{P(t,T_{i+1})}{P(t,T_{j+1})}$$

for $T_i < t \leq T_{i+1}$ and $T_j < T \leq T_{j+1}$.

1.2.4 Swap measures

Given a tenor structure $0 \leq T_0 < T_1 < \cdots < T_N$, for k,m satisfying $0 \leq k < N$ and $k + m \leq N$ we define:

(i) **Annuity factor**:

$$A_{k,m}(t) = \sum_{n=k}^{k+m-1} P(t,T_{n+1})\tau_n, \quad \tau_n = T_{n+1} - T_n$$

(ii) **Swap rate**:

$$S_{k,m}(t) = \frac{P(t,T_k) - P(t,T_{k+m})}{A_{k,m}(t)} = \frac{\sum_{n=k}^{k+m-1} \tau_n P(t,T_{n+1}) L_n(t)}{A_{k,m}(t)}, \quad t \le T_k$$

Being a combination of zero-coupon bonds, an annuity factor $A_{k,m}(t)$ qualifies as a numraire. Denoting by $Q^{k,m}$ the corresponding measure and by $E^{k,m}$ the corresponding expectation, in the absence of arbitrage we have

$$V(t) = A_{k,m} E^{k,m}\left[\frac{V(T)}{A_{k,m}(T)}\right]$$

1.3 The HJM analysis

The HJM framework attempts to model how an entire continuum of T-indexed bond prices $P(t,T)$ jointly evolves through time, staring from a known condition $P(0,T)$.

In the absence of arbitrage, the deflated bond values are Q-martingales, so if $P_\beta(t,T) = \frac{P(t,T)}{\beta(t)}$, by the martingale representation theorem there exists a d-dimensional stochastic process $\sigma_T(t,T)$ with $\sigma_P(T,T) = 0$ such that

$$dP_\beta(t,T) = -P_\beta(t,T)\sigma_P(t,T)^t dW(t), \quad t \le T$$

By It's lemma and the previous SDE it follows that

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

By It's lemma it also follows that the forward bond prices satisfy

$$\frac{dP(t,T,T+\tau)}{P(t,T,T+\tau)} = -[\sigma_P(t,T+\tau) - \sigma_P(t,T)]^t \sigma_P(t,T)dt - [\sigma_P(t,T+\tau) - \sigma_P(t,T)]^t dW(t)$$

In the T-forward measure, $P(t,T,T+\tau)$ is a martingale, so we must have

$$\frac{dP(t,T,T+\tau)}{P(t,T,T+\tau)} = -[\sigma_P(t,T+\tau) - \sigma_P(t,T)]^t dW^T(t)$$

where $W^T(t)$ is a Q^T-Brownian motion.

Comparison of the last two expressions yields

$$dW^T(t) = dW(t) + \sigma_P(t,T)dt$$

which by Girsanov's theorem identifies the Radon-Nikodym process for the measure shift

$$\xi(t) = E_t^Q\left[\frac{dQ^T}{dQ}\right] = \mathcal{E}_t(\sigma_P \cdot W) = \exp\left(-\int_0^t ||\sigma_P(s,T)||^2\, ds - \int_0^t \sigma_P(s,T)\, dW(s)\right)$$

HJM models are traditionally stated in terms of instantaneous forward rates $f(t,T)$. Once again by It's lemma we have (omitting the drift term)

$$d\ln P(t,T) = \mathcal{O}(dt) - \sigma_P(t,T)^t dW(t)$$

Differentiating with respect to T on both sides and recalling that $f(t,T) = -\frac{\partial P(t,T)}{\partial T}$ we have

$$df(t,T) = \mu_f(t,T) + \underbrace{\frac{\partial}{\partial T}\sigma_P(t,T)^t}_{:=\sigma_f(t,T)^t} dW(t)$$

Concretely, we have the following:

Lemma 1.1. *The process for $f(t,T)$ in the T-forward measure is*

$$df(t,T) = \sigma_f(t,T)^t dW^T(t)$$

and in the risk-neutral measure the process is

$$\begin{aligned} df(t,T) &= \sigma_f(t,T)^t \sigma_P(t,T)dt + \sigma_f(t,T)^t dW(t) \\ &= \sigma_f(t,T)^t \int_t^T \sigma_f(t,u)\, du dt + \sigma_f(t,T)^t dW(t) \end{aligned}$$

We thus have:

- The HJM model is fully specified once the forward rate diffusion coefficients $\sigma_f(t,T)$ have been specified for all t and T.
- The initial forward rates $f(0,T)$ are taken as exogenous inputs.
- The model suffers from an obvious dimensionality problem: in order to describe the time-t state of a discount bond curve spanning $[t,T]$ we need to keep track of a continuum of forward rates $\{f(t,u)\}_{t \leq u \leq T}$.

The short rate process in measure Q for the HJM model is

$$r(t) = f(t,t) = f(0,t) + \int_0^t \sigma_f(u,t)^t \int_u^t \sigma_f(u,s)\, ds\, du + \int_0^t \sigma_f(u,t)^t\, dW(u) \tag{1}$$

It is generally **not Markovian**: for the path-dependent term

$$D(t) := \int_0^t \sigma_f(u,t)^t\, dW(u)$$

we must have

$$D(T) = D(t) + \int_t^T \sigma_f(u,T)^\top\, dW(u) + \left[\int_0^t \sigma_f(u,T)\, dW(u) - \int_0^t \sigma_f(u,t)^\top\, dW(u) \right] \tag{2}$$

whereby

$$E^Q\left[D(T)|D(t)\right] \neq E_t^Q[D(T)]$$

unless the bracketed term in (2) is either non-random or a deterministic function of $D(t)$.

1.3.1 The Gaussian model

If $\sigma_P(t,T)$ is assumed to be a bounded d-dimensional deterministic function, then forward bond prices $P(t,T,T+\tau)$ are log-normally distributed and $r(T) = f(T,T)$ is Gaussian with Q-moments

$$E_t^Q[f(T,T)] = \int_t^T \sigma_f(u,T)^t \sigma_P(u,T)\, du, \quad Var_t^Q[f(T,T)] = \int_t^T \sigma_f(u,T)^t \sigma_f(u,T)\, du$$

Propositions 4.5.1, 4.5.2 and 4.5.3 in [AP10-1] illustrate how to easily price options on ZCB's, caplets and futures rates in the Gaussian HJM model.

1.3.2 Gaussian model with Markovian short rate

Caverhill showed in 1994 that the short rate process can be made Markovian by imposing a separability condition on the deterministic forward rate volatility function, namely

$$\sigma_f(t,T) = g(t)h(T) \tag{3}$$

where $h : \mathbb{R} \to \mathbb{R}$ is positive and $g : \mathbb{R} \to \mathbb{R}^{d \times 1}$. Using (3), the short rate expression (1) becomes

$$r(t) = f(0,t) + h(t) \int_0^t g(u)^\top g(u) \left(\int_u^t h(s)\, ds \right) du + \underbrace{h(t) \int_0^t g(u)^\top \, dW(u)}_{\stackrel{not}{=} D(t)} \tag{4}$$

and importantly, the term $D(t)$ is now Matkov since

$$D(T) = \frac{h(T)}{h(t)} D(t) + h(T) + \int_t^T g(u)^\top \, dW(u)$$

The following proposition shows that under the separability condition (3), the short rate process is the solution of an SDE, and is hence Markovian.

Proposition 1.2 (Markovian short rate under *separable* Gaussian HJM)**.** *In the d-dimensional Gaussian HJM model*[2] *and under the separability condition (3)*

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

the short rate process satisfies an SDE

$$dr(t) = (a(t) - \kappa(t) r(t))\, dt + \sigma_r(t)^\top dW(t)$$

where

$$\begin{aligned} h(T) &= e^{-\int_0^T \kappa(s)\, ds}, \\ g(t) &= e^{\int_0^t \kappa(s)\, ds} \sigma_r(t), \\ a(t) &= \frac{\partial f(0,t)}{\partial t} + \kappa(t) f(0,t) + \int_0^t \sigma_f(u,t)^\top \, du \end{aligned}$$

[2]Namely, under the assumption that $\sigma_P(t,T)$ is a bounded (d-dimensional) deterministic function.

2 Vanilla models with local volatility

Local or deterministic volatility models are one-factor diffusive models where our ability to alter the terminal distribution stems from a single source: a swap rate dependent function.

- Via copula methods, vanilla models can be extended to describe joint terminal distributions of more than one rate.
- Vanilla models serve as a foundation for the development of more widely applicable full term structure models.

2.1 General framework

Notation:

- $S(t)$: forward Libor rate
- P: measure under which $S(t)$ is a martingale satisfying the SDE

$$dS(t) = \lambda\varphi(S(t))dW(t), \qquad \lambda = \text{const}, \quad \varphi(0) = 0 \tag{5}$$

- $W(t)$: one-dimensional Brownian motion under P.

The **role of the function** φ is to match the distribution of S to that observed through calls and puts traded in the market. Remember that the time-t price of a European call struck at K maturing at T is given by rhe risk-neutral pricing formula

$$c(t, S(t); T, K) = E_t\left[(S(T) - K)^+\right]$$

The time-t **implied volatility** function $\sigma_B(t, S(t); T, K)$ is the implicit solution to the Black-Scholes formula

$$c(t, S; T, K) = S\mathcal{N}(d_+) - K\mathcal{N}(d_-), \quad d_\pm = \frac{\ln(S/K) \pm \frac{1}{2}\sigma_B^2(t, S; T, K)(T-t)}{\sigma_B(t, S; T, K)\sqrt{T-t}}$$

The mapping $T \mapsto \sigma_B(t, s; T, K)$ is the **T-maturity volatility smile**.

2.2 The CEV model

The **constant elasticity of variance (CEV) model** is a local volatility model (5) with $\varphi(S) = S^p$.

Proposition 2.1. *See [AP10-1, Proposition 7.2.1]. Consider the SDE*

$$\frac{\partial}{\partial S}(t) = \lambda S(t)^p dW(t), \quad p > 0 \tag{6}$$

1. *All solutions of (6) are non-explosive.*
2. *For $p \geq \frac{1}{2}$, the SDE (6) has a unique solution*[3].
3. *For $0 < p < 1$, $S = 0$ is an attainable boundary for (6); for $p \geq 1$, $S = 0$ is an unattainable boundary for (6).*
4. *For $0 < p \leq 1$, $S(t)$ in (6) is a martingale; for $p > 1$, $S(t)$ is a strict supermartingale.*

The transition probability of $S(t)$ in (6) is known in closed form (see Lemma 7.2.4 in [AP10-1]).

[3] Hence, if the solution ever reaches $S = 0$, it must stay there (i.e. it is **absorbed**). If $0 < p < \frac{1}{2}$, a boundary condition at $S = 0$ needs to be specified in order to get a unique solution.

2.2.1 Valuation of European call options in the CEV model

2.2.2 Regularization

The CEV process implies a positive probability of absorption at $S = 0$ for $p < 1$, which might create difficulties in pricing more exotic structures. **To avoid absorption**, we can specify a regularized version of the CEV model by letting

$$\varphi(x) = x \min\{\epsilon^{p-1}, x^{p-1}\}, \quad \epsilon > 1, \quad p < 1$$

With $\varphi(x)$ now being Lipschitz, the process $S(t)$ can no longer reach the origin.

- The above specification might not allow for closed-form option pricing.
- For moderate (small) values of ϵ, Andersen and Andreasen show that the formulas in subsection ?? provide good approximations (c.f. [AP10-1, Proposition 7.2.10]).

2.2.3 Displaced diffusion models

These are local volatility models (5) with $\varphi(S) = (\alpha + S)^p$, for some constant α.

Setting $Z(t) = \alpha + S(t)$, by It's lema we see that $Z(t)$ satisfies the CEV SDE

$$dZ(t) = \lambda[Z(t)]^p dW(t)$$

and call option pricing is hence straightforward:

Proposition 2.2. *Let*

$$c_{dCEV}(t, S(t); T, K, \alpha) = E_t\left[(S(T) - K)^+\right]$$

be the call option price associated with the displaced CEV process. Then

$$c_{dCEV}(t, S(t); T, K, \alpha) = c_{CEV}(t, S + \alpha; T, K + \alpha), \quad S, K > -\alpha$$

In practice, the main use of displacement constants is for the special case of the displaced log-normal process where $p = 1$.

Proposition 2.3. *Consider the displaced log-normal process*

$$dS(t) = \lambda(\beta + \xi S(t))dW(t), \quad \xi, \lambda \neq 0$$

where $W(t)$ is a 1-dimensional Brownian motion in measure P. Assuming $S(t), K > -\frac{\beta}{xi}$, we have

$$c_{DLN}(t, S(t); T, K) = E_t\left[(S(T) - K)^+\right] = \left(S(t) + \frac{\beta}{\xi}\right)\mathcal{N}(d_+) - \left(K + \frac{\beta}{\xi}\right)\mathcal{N}(d_-), \quad d_\pm = \cdots$$

Remark 2.1. A first-order approximation to any local volatility process is of displaced normal type. Indeed, given a general local volatility model (5)

$$dS(t) = \lambda\varphi(S(t))dW(t), \qquad \lambda = \text{const}, \quad \varphi(0) = 0,$$

expanding the local volatility function φ around at-the-money to first order, we obtain

$$dS(t) \simeq \lambda\varphi(S(0)) + \varphi'(S(0))(S(t) - S(0))dW(t)$$

which is of displaced log-normal type

$$dS(t) = \sigma(bS(t) + (1-b)L)dW(t), \quad \sigma = \lambda\frac{\varphi(S(0))}{S(0)}, \quad b = \varphi'(S(0))\frac{S(0)}{\varphi(S(0))}, \quad L = S(0)$$

2.3 Finite difference solutions

For general specifications of the function φ defining the local volatility model

$$dS(t) = \lambda\varphi(S(t))dW(t)$$

closed-form solutions for European option will not exist. One may instead try to obtain a numerical solution: for fixed T, K, the Feynman-Kac theorem translates the risk-neutral pricing formula

$$c(t, S(t)) = c(t, S(t); Y, K) = E_t\left[(S(T) - K)^+\right]$$

into a PDE

$$\frac{\partial c(t,S)}{\partial t} + \frac{1}{2}\lambda^2\varphi(S)^2\frac{\partial^2 c(t,S)}{\partial S^2} = 0, \qquad c(T,S) = (S-K)^+$$

whose solution can be approximated using finite difference methods.

2.3.1 Multiple λ and T

In applications we may **need to compute the values of** $c(t, S; T, K)$ **for many different values od** T **and/or** λ (e.g. within a standard calibration exercise, where a root-search algorithm is used to to determine the value of λ that minimizes the error between computed and market prices). Instead of solving the PDE multiple times, one may use the follwing result:

Proposition 2.4. *Let* $g(\tau, x)$ *be the solution to the PDE*

$$-\frac{\partial g(\tau,x)}{\partial\tau} + \frac{1}{2}\varphi(x)^2\frac{\partial^2 g(\tau,x)}{\partial x^2} = 0, \qquad g(0,x) = (x-K)^+ \tag{7}$$

If $c(t, S; \lambda)$ *denotes the solution to the backward PDE*

$$\frac{\partial c(t,S)}{\partial t} + \frac{1}{2}\lambda^2\varphi(S)^2\frac{\partial^2 c(t,S)}{\partial S^2} = 0, \qquad c(T,S) = (S-K)^+ \tag{8}$$

for a given value of λ *then*

$$c(t, S; T, K) = g(\lambda^2(T-t), S) \tag{9}$$

Hence, one may use finite differences construct a solution g to the PDE (7) on a (τ, S)-grid and then use (9) to recover $c(t, S; T, K)$ for arbitrary choices of S, λ, T by simple look-up or interpolation.

2.4 Forward equation for call options

Instead, we may want **to use different strikes for different values of** T. In order to accomplish this (again, without having to solve the entire PDE each time) we may employ **Dupire's forward equation**, in which calendar time t and the initial spot S are considered fixed, whereas maturity T and strike K are variable. As before, let $c(t, S(t); T, K) = E_t\left[(S(T) - K)^+\right]$ denote the price of a European call.

Proposition 2.5 (Dupire, c.g. Proposition 7.4.2 in [AP10-1])**.** *The function* $c(T, K) = c(t, S; T, K)$ *(with* t, S *fixed) satisfies the forward PDE*

$$-\frac{\partial c(T,K)}{\partial T} + \frac{1}{2}\lambda^2\varphi(K)^2\frac{\partial^2 c(T,K)}{\partial K^2} = 0, \qquad c(t,K) = (S-K)^+, \quad t > T \tag{10}$$

An analogous version of Proposition also holds in this setting.

2.5 Time-dependent local volatility function φ

So far we have assumed that $S(t)$ is driven by an SDE

$$dS(t) = \lambda\varphi(S(t))dW(t), \qquad \lambda = \text{const}, \quad \varphi(0) = 0 \tag{11}$$

and we will now be extending our treatment to models for which $\varphi(t, S(t))$ depends on t as well (not just through $S(t)$).

Even though vanilla models with time-dependent local volatility functions φ have limited use in fixed income modeling, they do arise as describing approximate dynamics of swap or Libor rates in term structure models.

2.5.1 Separable case

The case

$$dS(t) = \lambda(t)\varphi(S(t))dW(t)$$

is treated via a simple time change argument that reduces it to a model of the type (11).

Lemma 2.6. *See [AP10-1, Proposition 7.6.1]. Define*

$$\tau(t) = \int_0^t \lambda^2(u)\, du$$

and define $s(\cdot)$ by $S(t) = s(\tau(t))$, where $S(t)$ follows

$$dS(t) = \lambda(t)\varphi(S(t))dW(t)$$

Then

$$ds(\tau) = \varphi(s(\tau))d\tilde{W}(\tau), \quad s(0) = S(0)$$

where $\tilde{W}$ is Brownian motion.

The valuation problem

$$c(t, S(t); T, K) = E_t\left[(S(T) - K)^+\right]$$

can hence be rewritten as

$$c(\tau, s(\tau); T, K) = E\left[(s(\tau) - K)^+ | \tilde{\mathbb{F}}_{\tau(t)}\right]$$

which can be solved using the methods outlined earlier.

2.5.2 Skew averaging

In order to handle the general case

$$dS(t) = \varphi(t, S(t))dW(t), \qquad \lambda = \text{const}, \quad \varphi(0) = 0 \tag{12}$$

one can look for a time-independent local volatility function (a time average of the time-dependent function) that yields European option prices approximately matching prices from the time-dependent model. We first rewrite (12) as

$$dS(t) = \lambda(t)g(t, S(t))dW(t), \quad g(t, x) = \frac{\varphi(t, x)}{\varphi(t, S_0)}, \quad X_0 = S(0), \quad \lambda(t) = \varphi(t, X_0)$$

The idea is to fix $T > 0$ and to derive conditions that a time-independent function $\bar{g}(x)$ needs to satisfy so that

$$dS(t) = \lambda(t)g(t, S(t))dW(t)$$

can be replaced with

$$dS(t) = \lambda(t)\bar{g}(S(t))dW(t)$$

for the purposes of valuing T-expiry European options of all strikes.

For $\epsilon \geq 0$, define

$$g^\epsilon(t,x) = g(t, X_0 + \epsilon(x - X_0)), \quad \bar{g}^\epsilon(x) = \bar{g}(X_0 + \epsilon(x - X_0))$$

and define 2 sets of processes

$$\begin{aligned} dX^\epsilon(t) &= \lambda(t)g^\epsilon(t, X^\epsilon(t))dW(t), \quad X^\epsilon(0) = X_0 \\ dY^\epsilon(t) &= \lambda(t)\bar{g}^\epsilon(Y^\epsilon(t))dW(t), \quad Y^\epsilon(0) = X_0 \end{aligned}$$

The requirement that the distributions of X^ϵ and Y^ϵ be close can be formalized as finding

$$\bar{g}(\cdot) = \operatorname{argmin}\left\{\bar{g} : \quad q(\epsilon) := E\left[(X^\epsilon(T) - Y^\epsilon(T))^2\right]\right\}$$

Expanding

$$q(\epsilon) = q(0) + \epsilon q'(0) + \frac{1}{2}\epsilon^2 q''(0) + \mathcal{O}(\epsilon^3)$$

one shows that $q(0) = q'(0) = 0$ so minimizing $q(\epsilon)$ is equivalent to minimizing $q''(0)$. Then one has the following result:

Proposition 2.7. *See [AP10-1, Proposition 7.6.2]. Any function $\bar{g}$ that minimizes $q''(0)$ must satisfy the condition*

$$\frac{\partial g}{\partial x}\bar{g}(X_0) = \int_0^T \frac{\partial g}{\partial x}(t, X_0)\omega_T(t)\, dt \tag{13}$$

where

$$\omega_T(t) = \frac{v^2(t)\lambda^2(t)}{\int_0^T v^2(t)\lambda^2(t)\, dt}, \qquad v^2(t) = E\left[(X^0(0) - X_0)^2\right]$$

Typical examples of time-dependent local volatility functions $g(t,x)$, scaled to satisfy $g(t, S(0)) = 1$, are:

- Displaced log-normal function $g(t,x) = b(t)\frac{x}{S(0)} + (1 - b(t))$.
- CEV function $g(t,x) = \left(\frac{x}{S(0)}\right)^{p(t)}$.

The condition (13) does not define the function $\bar{g}$ uniquely, so one generally seeks to find a function $\bar{g}$ within the same family of functions they approximate. For instance, for the above examples we would look for functions

$$\bar{g}(x) = \bar{b}\frac{x}{S(0)} + (1 - \bar{b}), \quad \bar{g}(x) = \left(\frac{x}{S(0)}\right)^{\bar{p}}$$

Concretely, we have the following useful result:

Corollary 2.8. *Over the time horizon $[0,T]$:*

1. *The effective skew $\bar{b}$ in $\bar{g}(x) = \bar{b}\frac{x}{S(0)} + (1-\bar{b})$ for the model defined by the time-dependent local volatility function $g(t,x) = b(t)\frac{x}{S(0)} + (1-b(t))$ is given by*

$$\bar{b} = \int_0^T b(t)\omega_T(t)\,dt$$

where

$$\begin{aligned}\omega_T(t) &= \frac{v(t)^2\lambda(t)^2}{\int_0^T v(t)^2\lambda(t)^2\,dt},\\ v(t) &= \int_0^t \lambda(s)^2\,ds\end{aligned}$$

2. *Similarly, the effective parameter $\bar{p}$ in $\bar{g}(x) = \left(\frac{x}{S(0)}\right)^{\bar{p}}$ for the model defined by the time-dependent local volatility function $g(t,x) = \left(\frac{x}{S(0)}\right)^{p(t)}$ is given by*

$$\bar{p} = \int_0^T p(t)\omega_T(t)\,dt$$

with $\omega_T(t)$ and $v(t)$ as above.

3. *If, furthermore, the volatility is constant $\lambda(t) \equiv \lambda$, we obtain*

$$v(t)^2 = \lambda^2 t, \qquad \omega_T(t) = \frac{t}{T^2/2}$$

3 Vanilla models with stochastic volatility

Here we focus on models which incorporate a new stochastic component for the volatility:

$$\begin{aligned} dS(t) &= \lambda\left(bS(t)+(1-b)L\right)\sqrt{z(t)}dW(t) && (14)\\ dz(t) &= \theta(z_0-z(t))dt+\eta\sqrt{z(t)}dZ(t) && (15)\\ E[dW(t)dZ(t)] &= \rho dt \end{aligned}$$

3.1 Basic properties

Proposition 3.1 (Explosions).

Proposition 3.2 (Martingale property).

The stochastic equation (14) can be integrated explicitly:

Proposition 3.3. *In the model (14)-(15), we have*

$$S(t)=\frac{1}{b}\left[(bS(0)+(1-b)L)X(t)-(1-b)L\right]$$

where

$$\ln X(t)=\lambda b\int_0^t\sqrt{z(s)}dW(s)-\frac{1}{2}\lambda^2b^2\int_0^t z(s)\,ds \tag{16}$$

As we will see, the moment generating function $\ln X(t)$ is of fundamental importance for European option pricing:

Proposition 3.4. *Consider the moment generating function of* $\ln X(t)$

$$\Psi_X(u;t)\stackrel{def}{=}E[e^{u\ln X(t)}]$$

In the model (14)-(15), for any $u\in\mathbb{C}$ *we have*

$$\Psi_X(u;t)=\Psi_{\bar{z}}\left(\frac{1}{2}(\lambda b)^2u(u-1),u;t\right)$$

where

$$\Psi_{\bar{z}}(v,u;t)=E^{\tilde{P}}\left[e^{v\bar{z}(t)}\right],\qquad \bar{z}(t)=\int_0^t z(s)\,ds$$

Here $\tilde{P}$ *is the measure defined by* $\mathcal{E}_t\left(u\lambda b\sqrt{z}\cdot W\right)$.

3.2 Fourier integration

3.2.1 Calculation of option prices when MGF is available for $\ln(S(t))$.

Theorem 3.5. *Let* ξ *be a random variable with moment generating function (MGF)* $\chi(u)=E[e^{u\xi}]$. *For* $k\in\mathbb{R}$ *and any*[4] $0<\alpha<1$ *we have*

$$\begin{aligned} E\left[(e^\xi-e^k)^+\right] &= \chi(1)-\frac{e^k}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-k(\alpha+iw)}\chi(\alpha+iw)}{(\alpha+iw)(1-\alpha-iw)}\,dw && (17)\\ &= \chi(1)-\frac{e^k}{\pi}\int_0^{\infty}Re\left[\frac{e^{-k(\alpha+iw)}\chi(\alpha+iw)}{(\alpha+iw)(1-\alpha-iw)}\right]dw && (18) \end{aligned}$$

[4]Here α is a dampening constant which is introduced in the proof as a factor $e^{-\alpha k}$ in order to guarantee that the Fourier transform exists.

Combining Theorem 3.5 (using a damping factor $\alpha = \frac{1}{2}$) with the closed-form expression for the moment generating function in the SV model given in Proposition 3.4, we obtain an efficient formula for pricing European call and put options in the SV model (14)-(15).

Theorem 3.6. *The price of a call option $c_{SV}(0, S; T, K)$ in the SV model (14)-(15) is given by*

$$c_{SV}(0, S; T, K) = \frac{1}{b} c_B(0, S'; T, K', \lambda b) - \frac{K'}{2\pi b} \int_{-\infty}^{\infty} \frac{e^{(1/2+iw)\ln(S'/K')} q\left(\frac{1}{2} + iw\right)}{w^2 + \frac{1}{4}} dw \tag{19}$$

where $c_B(0, S'; T, K', \sigma)$ is the Black formula for spot S', strike K', expiry T and volatility σ, with

$$S' = bS + (1-b)L, \quad K' = bK + (1-b)L$$

and

$$q(u) = \Psi_{\bar{Z}}\left(\frac{1}{2}(\lambda b)^2 u(u-1), u; T\right) - e^{\frac{1}{2}\lambda^2 b^2 z_0 T u(u-1)}$$

where $\Psi_{\bar{z}}$ is given in Proposition 3.4.

3.2.2 Direct integration

In order to compute the integral inside equation (19)

$$\int_{-\infty}^{\infty} \frac{e^{(1/2+iw)\ln(S'/K')} q\left(\frac{1}{2} + iw\right)}{w^2 + \frac{1}{4}} dw$$

we first need to truncate the domain.

Lemma 3.7. *Given that $|\rho| < 1$, the function*

$$q(u) = \Psi_{\bar{Z}}\left(\frac{1}{2}(\lambda b)^2 u(u-1), u; T\right) - e^{\frac{1}{2}\lambda^2 b^2 z_0 T u(u-1)}$$

defined in Theorem 3.6 satisfies

$$\lim_{w \to +\infty} \frac{1}{w} \ln\left[q(\frac{1}{2} + iw)\right] = -q_\infty$$

where

$$q_\infty = \frac{\lambda b z_0}{\eta}\left(\sqrt{1-\rho^2} + i\rho\right)(1 + \theta T)$$

One then shows easily that

$$\left|\int_{w_{max}}^{\infty} \frac{e^{(1/2+iw)\ln(S'/K')} q(\frac{1}{2} + iw}{w^2 + \frac{1}{4}} dw\right| \leq \sqrt{\frac{S'}{K'}} \frac{e^{-\mathrm{Re}(q_\infty) w_{\max}}}{w_{\max}}$$

so if ϵ_w is the absolute tolerance (generally $\epsilon = 10^{-3}, 10^{-6}$) one can set the upper truncation limit w_{max} by the condition

$$\frac{e^{-\mathrm{Re}(q_\infty) w_{\max}}}{w_{\max}} = \epsilon_w$$

We can then apply anu quadrature rule in order to discretize the integral. For instance, the trapezoidal rule yields

$$\ldots$$

3.2.3 Fourier integration for arbitrary European Payoffs

Given a payoff function $f(x)$, recall that the value of a European option with payoff $f(S(T))$ is given by[5]

$$E[f(S(T))] = \int f(x)\varphi_{S(T)}(x)\,dx = \int f(x)\frac{\partial^2 c(0,S(0);T,x)}{\partial x^2}\,dx$$

where $\varphi_{S(T)}(x)$ is the density function of $S(T)$ and $c(0,S(0);T,K)$ is the European call option value for $S(\cdot)$. Integrating by parts we obtain a useful representation of a general European payoff in terms of European calls and puts.

Proposition 3.8. *For any twice-continuously differentiable function $f(x)$, the value of a European option with payoff $f(\cdot)$ maturing at T equals the weighted integral*

$$E[f(S(T))] = f(K^*) + f'(K^*)(S(0)-K^*) + \int_{-\infty}^{K^*} p(0,S(0);T,K)f''(K)\,dK + \int_{K^*}^{\infty} c(0,S(0);T,K)f''(K)\,dK \quad (20)$$

for any K^*.

Hence, in order to compute the value of a European option with arbitrary payoff we need to simultaneously compute call option prices for different strikes, and the **FFT method** provides an efficient way of accomplishing this. The integrals we need to evaluate are

$$I(K') = \int_{-\infty}^{\infty} \frac{e^{-(1/2+iw)\ln(S'/K')}}{w^2+\frac{1}{4}} q\left(\frac{1}{2}+iw\right)\,dw$$

for various K', where recall

$$S' = bS + (1-b)L, \quad K' = bK + (1-b)L$$

We discretize K' in such a way that $\ln(\frac{S'}{K'})$ are equidistant. In particular, choosing step $\delta > 0$ we define

$$x_n = \delta n, \qquad K'_n = S'e^{x_n}, \qquad n = -N,\dots,N$$

so that

$$\ln\left(\frac{S'}{K'_n}\right) = -x_n, \quad K_n = \left(S + \frac{1-b}{b}L\right)e^{x_n} - \frac{1-b}{b}L$$

and then

$$\begin{aligned} I_n &\overset{not}{=} I(K'_n) = \int_{-\infty}^{\infty} \frac{e^{-(1/2+iw)\delta n}}{w^2+\frac{1}{4}} q\left(\frac{1}{2}+iw\right)\,dw = e^{-0.5\delta n} J_n \\ J_n &= \int_{-\infty}^{\infty} e^{-iw\delta n}\frac{q\left(\frac{1}{2}+iw\right)}{w^2+\frac{1}{4}}\,dw \end{aligned}$$

Once this is accomplished, the value of an option with general payoff $f(\cdot)$ is obtained by discretizing the integrals in equation (20).

[5] The price of a call option is given by $C(S,K) = \int_K^\infty \varphi_{S(T)}(x)(x-K)\,dx$ Differentiating with respect to K twice we obtain

$$\begin{aligned} \frac{\partial C(S,K)}{\partial K} &= \frac{\partial}{\partial K}\left[\int_K^\infty \varphi_{S(T)}(x)\,dx - K\int_K^\infty \varphi_{S(T)}(x)\,dx\right] = -\int_K^\infty \varphi_{S(T)}(x)\,dx \\ \frac{\partial^2 C(S,K)}{\partial K^2} &= \varphi_{S(T)}(K) \end{aligned}$$

Proposition 3.9. *Fix $\delta > 0$. Let K_n, K'_n be defined by*

$$K_n = \left(S + \frac{1-b}{b}L\right)e^{x_n} - \frac{1-b}{b}L, \quad K'_n = S'e^{x_n}, \qquad n = -N, \dots, N$$

The value of a call option with payoff $f(\cdot)$ at time T in the SV model (14)-(15) is approximately given by

$$E[f(S(T))] = f(S(0)) + \sum_{n=-N}^{-1} p_{SV}(0, S(0); T, K_n) f''(K_n)(K_{n+1} - K_n) \\ + \sum_{n=0}^{N-1} c_{SV}(0, S(0); T, K_n) f''(K_n)(K_{n+1} - K_n)$$

where

$$\begin{aligned} c_{SV}(0, S(0); T, K_n) &= \frac{1}{b} c_B(0, S'; T, K'_n; \lambda b) - \frac{K'_n}{2\pi b} e^{-0.5\delta n} J_n \\ p_{SV}(0, S(0); T, K_n) &= \frac{1}{b} p_B(0, S'; T, K'_n; \lambda b) - \frac{K'_n}{2\pi b} e^{-0.5\delta n} J_n \end{aligned}$$

with $\{J_n\}_{n=-N}^{N}$ evaluated by an inverse FFT transform of the function

$$\frac{q\left(\frac{1}{2} + iw\right)}{w^2 + \frac{1}{4}}, \quad q(u) = \Psi_{\bar{Z}}\left(\frac{1}{2}(\lambda b)^2 u(u-1), u; T\right) - e^{\frac{1}{2}\lambda^2 b^2 z_0 T u(u-1)}$$

Remark 3.1 (FFT reminder)**.** The discrete Fourier transform maps

$$\mathbf{x} = (x_1, \dots, x_N) \mapsto \hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_N), \qquad \left\{ \begin{array}{l} \hat{x}_k = \sum_{j=1}^{N} e^{-i\frac{2\pi}{N}(j-1)(k-1)} x_j \\ x_k = \frac{1}{N} \sum_{j=1}^{N} e^{i\frac{2\pi}{N}(j-1)(k-1)} \hat{x}_j \end{array} \right. \tag{21}$$

Computing these sums independently would require N^2 operations but the FFT algorithm accomplishes this with only $N \log_2(N)$ operations. The idea is to reduce the computation of an integral to the computation of an inverse Fourier transform to which the FFT algorithm can be applied. We seek to compute the integrals

$$J_n = 2\int_0^\infty \mathrm{Re}\left[e^{-iw\delta n} \frac{q\left(\frac{1}{2} + iw\right)}{w^2 + \frac{1}{4}}\right] dw, \qquad n = -N, \cdots, N$$

In order to do so we truncate the domain of integration to $[0, w_{max}]$ according to the discussion in section 4.2.2 and we discretize the integral using, for instance, the trapezoidal rule

$$\int_a^b f(w)\, dw = \sum_{k=0}^{N_w} \alpha_k f(w_k), \quad w_k = k\eta, \quad \eta = \frac{b-a}{N_w}, \quad \alpha_k = \left\{ \begin{array}{ll} 1/2, & k = 0, N \\ 1, & 2 \le k \le N-1 \end{array} \right.$$

Our discretized integral hence becomes

$$J_n \simeq 2\sum_{k=0}^{N_w} \alpha_k \mathrm{Re}\left[e^{-iw_k\delta n} \frac{q\left(\frac{1}{2} + iw_k\right)}{w_k^2 + \frac{1}{4}}\right] = \sum_{k=0}^{N_w} \alpha_k \mathrm{Re}\left[e^{-i\eta\delta kn} \frac{q\left(\frac{1}{2} + ik\eta\right)}{(k\eta)^2 + \frac{1}{4}}\right], \quad n = -N, \cdots, N$$

and choosing η such that $\eta\delta = \frac{2\pi}{N_w}$ our integrals become

$$J_n \simeq 2\sum_{k=0}^{N_w} \alpha_k \mathrm{Re}\left[e^{-i\frac{2\pi}{N_w}kn}\frac{q\left(\frac{1}{2}+ik\eta\right)}{(k\eta)^2+\frac{1}{4}}\right], \quad n = -N, \cdots, N$$

and from (21) we can identify the J_n's as the inverse Fourier transforms of the x_k's where $x_k = \frac{q\left(\frac{1}{2}+ik\eta\right)}{(k\eta)^2+\frac{1}{4}}$.

3.3 Averaging methods for models with time-dependent parameters

Models with time-dependent parameters emerge naturally when vanilla models are used to approximate interest rate dynamics in a full term structure model. Our goal here is two-fold:

1. Given a stochastic volatility model with time dependent parameters

$$\begin{aligned} dS(t) &= \lambda(t)\left[b(t)S(t) + (1-b(t))L\right]\sqrt{z(t)}dW(t) && (22)\\ dz(t) &= \theta(z_0 - z(t))dt + \eta(t)\sqrt{z(t)}dZ(t) && (23)\\ E[dW(t)dZ(t)] &= \rho dt \end{aligned}$$

   we find a constant time-averaged volatility $\bar\lambda$, skew $\bar b$ and volatility of variance $\bar\eta$ such that the model

$$\begin{aligned} dS(t) &= \bar\lambda\left[\bar bS(t) + (1-\bar b)L\right]\sqrt{z(t)}dW(t) && (24)\\ dz(t) &= \theta(z_0 - z(t))dt + \bar\eta\sqrt{z(t)}dZ(t) && (25)\\ E[dW(t)dZ(t)] &= \rho dt \end{aligned}$$

   produces the same European option prices as the model (22)-(23).
2. We illustrate how to employ time-averaging techniques to calibrate the SV model with time-dependent parameters (22)-(23) to quoted market prices.

3.3.1 Volatility averaging

We start by letting the volatility $\lambda(t)$ be time-dependent, while holding the skew b and the volatility of variance η constant.

Theorem 3.10 (effective volatility approximation)**.** *Values of European options with expiry T in the model*

$$\begin{aligned} dS(t) &= \lambda(t)\left[bS(t) + (1-b)L\right]\sqrt{z(t)}dW(t)\\ dz(t) &= \theta(z_0 - z(t))dt + \eta\sqrt{z(t)}dZ(t)\\ E[dW(t)dZ(t)] &= \rho dt \end{aligned}$$

are well approximated by their values in the model

$$\begin{aligned} dS(t) &= \bar\lambda\left[bS(t) + (1-b)L\right]\sqrt{z(t)}dW(t)\\ dz(t) &= \theta(z_0 - z(t))dt + \eta\sqrt{z(t)}dZ(t)\\ E[dW(t)dZ(t)] &= \rho dt \end{aligned}$$

where the effective SV volatility $\bar\lambda$ solves the equation

$$\Psi_{\bar z}\left(\frac{h''(\zeta_T)}{h'(\zeta_T)}\bar\lambda^2, 0; T\right) = \Psi_{\overline{z\lambda^2}}\left(\frac{h''(\zeta_T)}{h'(\zeta_T)}\bar\lambda^2, 0; T\right) \quad (26)$$

where

$$\zeta_T = z_0 \int_0^T \lambda(t)^2\,dt, \qquad h(x) = \frac{bS(0)+(1-b)L}{b}\left[2\mathcal{N}\left(\frac{b\sqrt{x}}{2}\right)-1\right]$$

and where $\Psi_{\bar{z}}$ and $\Psi_{\overline{z\lambda^2}}$ are moment-generating functions given by [AP10-1, Proposition 8.3.8] and [AP10-1, Proposition 9.1.2] respectively (see remark below for details).

Remark 3.2. The LHS of equation (26) can be computed in closed form via [AP10-1, Proposition 8.3.8]:

$$\begin{aligned}
\Psi_{\bar{z}}(v,u;T) &= e^{A(v,u)+B(v,u)z_0},\\
A(v,u) &= \frac{\theta z_0}{\eta^2}\left[2\ln\left(\frac{2\gamma}{\theta'+\gamma-e^{-\gamma T}(\theta'-\gamma)}\right)+(\theta'-\gamma)T\right],\\
B(v,u) &= \frac{2v(1-e^{-\gamma T})}{(\theta'+\gamma)(1-e^{-\gamma T})+2\gamma e^{-\gamma T}},\\
\gamma=\gamma(v,u) &= \sqrt{(\theta')^2-2\eta^2 v},\\
\theta'=\theta'(u) &= \theta-\rho\eta\lambda b u
\end{aligned}$$

As for the RHS, assuming that $\lambda(t)$ is piecewise constant $\lambda(t)=\sum_{i=1}^n \lambda_i \mathbb{I}_{(t_{i-1},t_i]}(t)$ for some $0=t_0<\cdots<T_n=T$, then on each interval $(t_{i-1},t_i]$ we can apply the closed-form solution from the previous paragraph, and piece all these together.

Equation (26) can then be solved in a couple of iterations of the Netwon-Raphson method.

3.3.2 Skew averaging

We now also allow the skew b to be time-dependent and we find an effective skew $\bar{b}$.

Proposition 3.11 (effective skew approximation)**.** *Values of European options with expiry T in the model*

$$\begin{aligned}
dS(t) &= \lambda(t)\left[b(t)S(t)+(1-b(t))L\right]\sqrt{z(t)}dW(t)\\
dz(t) &= \theta(z_0-z(t))dt+\eta\sqrt{z(t)}dZ(t)\\
E[dW(t)dZ(t)] &= \rho dt
\end{aligned}$$

are well approximated by their values in the model

$$\begin{aligned}
dS(t) &= \lambda(t)\left[\bar{b}S(t)+(1-\bar{b})L\right]\sqrt{z(t)}dW(t)\\
dz(t) &= \theta(z_0-z(t))dt+\eta\sqrt{z(t)}dZ(t)\\
E[dW(t)dZ(t)] &= \rho dt
\end{aligned}$$

where the effective SV skew $\bar{b}$ is given by

$$\bar{b} = \int_0^T b(t)\omega_T(t)\,dt$$

where the weights are given by

$$\begin{aligned}
\omega_T(t) &= \frac{v(t)^2\lambda(t)^2}{\int_0^T v(t)^2\lambda(t)^2\,dt},\\
v(t)^2 &= z_0^2\int_0^t \lambda(s)^2\,ds + z_0\eta^2 e^{-\theta t}\int_0^t \lambda(s)^2\frac{e^{\theta s}-e^{-\theta s}}{2\theta}\,ds
\end{aligned}$$

3.3.3 Volatility of variance averaging

Finally we also allow the volatility of variance to be time-dependent.

Proposition 3.12 (effective volatility of variance approximation)**.** *Values of European options with expiry T in the model*

$$
\begin{aligned}
dS(t) &= \lambda(t)\left[b(t)S(t) + (1-b(t))L\right]\sqrt{z(t)}dW(t) \\
dz(t) &= \theta(z_0 - z(t))dt + \eta(t)\sqrt{z(t)}dZ(t) \\
E[dW(t)dZ(t)] &= \rho dt
\end{aligned}
$$

are well approximated by their values in the model

$$
\begin{aligned}
dS(t) &= \lambda(t)\left[b(t)S(t) + (1-b(t))L\right]\sqrt{z(t)}dW(t) \\
dz(t) &= \theta(z_0 - z(t))dt + \bar{\eta}\sqrt{z(t)}dZ(t) \\
E[dW(t)dZ(t)] &= \rho dt
\end{aligned}
$$

where the effective SV volatility of variance $\bar{\eta}$ is given by

$$
\bar{\eta}^2 = \frac{\int_0^T \eta(t)^2 \rho_T(t)\, dt}{\int_0^T \rho_T(t)\, dt}
$$

where the weight function is given by

$$
\rho_T(u) = \int_u^T ds \int_s^T dt \lambda(t)^2 \lambda(s)^2 e^{-\theta(t-s)} e^{-2\theta(s-r)}.
$$

3.3.4 Calibration by parameter averaging

Given:

- A collection of expiries is given $0 = T_0 < \cdots < T_N$.
- A collection of strikes $K_1, \ldots, K_M$.
- Market values of European call options with expiries T_n and strikes K_m, namely

$$
\{\hat{c}_{n,m}, \quad n = 1, \ldots, N, \quad m = 1, \ldots, M\}
$$

The goal is to find time-dependent model parameters $\lambda(t)$, $b(t)$ and $\eta(t)$ such that the model

$$
\begin{aligned}
dS(t) &= \lambda(t)\left[b(t)S(t) + (1-b(t))L\right]\sqrt{z(t)}dW(t) \\
dz(t) &= \theta(z_0 - z(t))dt + \eta(t)\sqrt{z(t)}dZ(t) \\
E[dW(t)dZ(t)] &= \rho dt
\end{aligned}
$$

values the above options as close as possible. For instance, if $\mathcal{X} = \{\lambda(\cdot), b(\cdot), \eta(\cdot)\}$ denotes the state of the model and $c_{n,m}(\mathcal{X})$ are the model option prices, we seek to find

$$
\{\lambda(\cdot), b(\cdot), \eta(\cdot)\} \text{arg min} \sum_{n,m} \left(c_{n,m}(\mathcal{X}) - \hat{c}_{n,m}\right)^2
$$

Alternatively, we can work with market parameter values: denote by $\{\hat{\lambda}, \hat{b}, \hat{\eta}\}$, $n = 1, \ldots, N$ determined in such a way that the market prices of European options expiring at time T_n, namely $\{\hat{c}_{n,m}\}_{m=1}^M$ match the prices obtained in the model

$$
\begin{aligned}
dS(t) &= \hat{\lambda}_n \left[\hat{b}_n S(t) + (1-\hat{b}_n)L\right]\sqrt{z(t)}dW(t) \\
dz(t) &= \theta(z_0 - z(t))dt + \hat{\eta}_n\sqrt{z(t)}dZ(t)
\end{aligned}
$$

Denote by $\{\lambda(\mathcal{X}), b(\mathcal{X}), \eta(\mathcal{X})\}$ the averaged parameters to time T_n for the time-dependent model (the results in the previous section relate these to $\{\hat{\lambda}, \hat{b}, \hat{\eta}\}$). We can then instead consider the following more convenient minimization problem

$$\{\lambda(\cdot), b(\cdot), \eta(\cdot)\} = \arg\min \Bigg\{ W_\lambda \sum_{n,m} \left(\bar{\lambda}_n(\mathcal{X}) - \hat{\lambda}_n\right)^2 + W_b \sum_{n,m} \left(\bar{b}_n(\mathcal{X}) - \hat{b}_n\right)^2 + W_\eta \sum_{n,m} \left(\bar{\eta}_n(\mathcal{X}) - \hat{\eta}_n\right)^2 \Bigg\}$$

with appropriate user-specified weights W_λ, W_b and W_η.

Calibration can then be split into independent sub-calibrations as follows. Recall the three averaging results we outlined in the previous section:

$$\begin{aligned}
\bar{\eta}_n(\mathcal{X})^2 &= \frac{\int_0^{T_n} \eta(t)^2 \rho_{T_n}(t, \lambda(\cdot))\, dt}{\int_0^{T_n} \rho_{T_n}(t, \lambda(\cdot))\, dt}, \qquad n = 1, \ldots, N \\
\bar{b}_n(\mathcal{X}) &= \int_0^{T_n} b(t)\omega_{T_n}(t; \lambda(\cdot), \eta(\cdot))\, dt, \qquad n = 1, \ldots, N \\
\bar{\lambda}_n(\mathcal{X}) &= F\left(\lambda(\cdot), \bar{b}_n(\mathcal{X}), \bar{\eta}_n(\mathcal{X})\right), \qquad \text{where} \\
F(\lambda(\cdot); \bar{b}_n(\mathcal{X}), \bar{\eta}_n(\mathcal{X})) &= \sqrt{\frac{h'(\zeta_{T_n})}{h''(\zeta_{T_n})} \cdot \Psi_{\bar{\bar{z}}}^{-1}\left(\Psi_{\overline{z\lambda^2}}\left(\frac{h'(\zeta_{T_n})}{h''(\zeta_{T_n})}, 0; T\right)\right), 0; T}
\end{aligned}$$

Assume the model parameters are piecewise constant between option expiry dates:

$$\lambda(t) = \sum_{i=1}^{N} \lambda_i \mathbb{I}_{(T_{i-1}, T_i]}(t), \quad b(t) = \sum_{i=1}^{N} b_i \mathbb{I}_{(T_{i-1}, T_i]}(t), \quad \eta(t) = \sum_{i=1}^{N} \eta_i \mathbb{I}_{(T_{i-1}, T_i]}(t)$$

We can then also discretize the weights as

$$\begin{aligned}
\rho_{T_n}(t, \lambda(\cdot)) &= \sum_{i=1}^{n} \rho_{n,i}(\lambda(\cdot)) \mathbb{I}_{(T_{i-1}, T_i]}(t), \\
\omega_{T_n}(t; \lambda(\cdot), \eta(\cdot)) &= \sum_{i=1}^{n} \omega_{n,i}(\lambda(\cdot), \eta(\cdot)) \mathbb{I}_{(T_{i-1}, T_i]}(t)
\end{aligned}$$

and denote

$$\bar{\rho}_{n,i}(\lambda(\cdot)) = \frac{\rho_{n,i}(\lambda(\cdot))}{\int_0^{T_n} \rho_{T_n}(y; \lambda(\cdot))\, dt}$$

We are thus reduced to solving the three systems of equations

$$\sum_{i=1}^{n} \bar{\rho}_{n,i}(\lambda(\cdot))(T_i - T_{i-1})\eta_i^2 = \hat{\eta}_n^2, \tag{27}$$

$$\sum_{i=1}^{n} \omega_{n,i}(\lambda(\cdot), \eta(\cdot))(T_i - T_{i-1}) b_i = \hat{b}_n, \tag{28}$$

$$F\left(\lambda(\cdot); \bar{b}_n(\mathcal{X}), \bar{\eta}_n(\mathcal{X})\right) = \hat{\lambda}_n \tag{29}$$

for $n = 1, \ldots, N$.

This system can be solved iteratively as follows:

(i) We first solve equations (29) by replacing the model parameters $\hat{b}_n(\mathcal{X}), \bar{\eta}_n(\mathcal{X})$ with their market values, namely

$$F\left(\lambda(\cdot); \hat{b}, \hat{\eta}\right) = F\left(\lambda_1, \ldots, \lambda_n; \hat{b}, \hat{\eta}\right) = \hat{\lambda}_n, \quad n = 1, \ldots, N$$

which is a system of N decoupled one-dimensional equations. For instance, the n-th equation reads

$$F\left(\lambda_1^*, \ldots, \lambda_{n-1}^*, \lambda_n; \hat{b}, \hat{\eta}\right) = \hat{\lambda}_n$$

where λ_i^* denote model parameters already solved for.

(ii) We next solve the system (27) for η_i^2, $i = 1, \ldots, N$, using the values of λ_i found in (i), namely

$$\sum_{i=1}^{n} \bar{\rho}_{n,i}(\lambda^*(\cdot))(T_i - T_{i-1})\eta_i^2 = \hat{\eta}_n^2, \quad n = 1, \ldots, N$$

This also be done sequentially.

(iii) Finally we solve the linear system

$$\sum_{i=1}^{n} \omega_{n,i}(\lambda^*(\cdot), \eta^*(\cdot))(T_i - T_{i-1})b_i = \hat{b}_n, \quad n = 1, \ldots, N$$

for b_i, again sequentially.

4 One-factor short rate models

The general HJM class with its infinite-dimensional Markovian dynamics is to unwieldy to work in practice, so we seek to identify HJM model subclasses that involve a finite number of Markovian state variables only.

We start by reviewing the notions of swaps and swaptions. We then outline the most basic one-factor models and show how to calibrate them to current prices of ZCB's and swaptions. We also illustrate how to use these models to price European derivatives.

4.1 Review of Swaps and swaptions

A *swap* is a generic term for an OTC derivative in which two counterparties agree to exchange one stream (leg) of cash flows against another stream. A *fixed-for-floating interest rate swap* is a swap in which one leg is a stream of fixed rate payments and the other is a stream of payments based on a floating rate, generally Libor.

Consider an increasing sequence of maturity times (the tenor structure)

$$0 \leq T_0 < T_1 < \cdots < T_N, \qquad \tau_n = T_{n+1} - T_n$$

From the perspective of the fixed rate payer, the net cash flow of the swap at time T_{n+1}, per unit of notional, is

$$\tau_n(L_n(T_n) - k), \qquad L_n(t) = L(t, T_n, T_{n+1})$$

By the fundamental valuation result, the value os a swap is hence equal to

$$\begin{aligned}
V_{swap}(t) &= \beta(t) \sum_{n=1}^{N-1} \tau_n E_t \left[\frac{1}{\beta(T_{n+1})} (L_n(T_n) - k) \right] \\
&= \sum_{n=0}^{N-1} \left[P(t, T_n) - P(t, T_{n+1}) - \tau_n k P(t, T_{n+1}) \right] \\
&= \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1})(L_n(t) - k) \\
&= A(t) \left[S(t) - k \right]
\end{aligned}$$

where

$$A(t) =, \qquad S(t) =$$

are the annuity and the forward rate o the swap, respectively. We hence observe that a vanilla fixed-floating swap can be valued at time t using only the term structure of interest rates observed on that date.

A *European swaption* is an option giving the holder the right, but not he obligation, to enter a swap at a future date at a given fixed rate. Assuming the underlying swap starts on the expiry date T_0 of the option, for a payer swap we have

$$\begin{aligned}
V_{swaption}(T_0) &= \left[V_{swap}(T_0) \right]^+ = \left(\sum_{n=0}^{N-1} \tau_n P(T_0, T_{n+1})(L_n(T_0) - k) \right)^+ \\
V_{swaption}(t) &= \beta(t) E_t \left[\frac{1}{\beta(T_0)} V_{swaption}(T_0) \right] \\
&= \beta(t) \left(\frac{1}{\beta(T_0)} \sum_{n=0}^{N-1} \tau_n P(T_0, T_{n+1})(L_n(T_0) - k) \right)^+
\end{aligned}$$

4.2 The Ho-Lee model

Assume that the short rate follows an SDE

$$dr(t) = \sigma_r dW(t) \Longrightarrow r(t) = r(0) + \sigma_r W(t)$$

With only 2 parameters $r(0)$ and σ_r it is not possible to match the observable discount bond prices, so instead we may consider

$$r(t) = r(0) + a(t) + \sigma_r W(t), \quad a(0) = 0$$

Choosing $a(t)$ accordingly, we can match the discount bond curve at time 0:

Lemma 4.1 (definition of the Ho-Lee model)**.** *Assume*

$$\begin{aligned} r(t) &= r(0) + a(t) + \sigma_r W(t), \\ a(t) &= f(0,t) - r(0) + \frac{1}{2}\sigma_r^2 t^2. \qquad \textit{[Ho-Lee model]} \\ f(0,t) &= -\frac{\partial f(0,t)}{\partial t} \end{aligned}$$

Then for any $T > 0$ we have

$$E_t\left[e^{-\int_0^T r(u)\,du}\right] = P(0,T)$$

This model can be characterized further:

Lemma 4.2 (characterization of the Ho-Lee model)**.** *In the Ho-Lee model, the risk-neutral process for $r(t)$ and the time-t bond prices are respectively given by*

$$\begin{aligned} dr(t) &= \left(\partial_t f(0,t) + \sigma_r^2 t\right) dt + \sigma_r dW(t) \\ P(t,T) &= \frac{P(0,T)}{P(0,t)} \exp\left(-(r(t) - f(0,t))(T-t) - \frac{1}{2}\sigma_r^t t(T-t)^2\right) \end{aligned}$$

A few observations are in order:

- The Ho-Lee model could have been specified as an HJM model with constant forward rate volatility

$$\sigma_f(t,T) = \sigma_r, \qquad \sigma_P(t,T) = \int_t^T \sigma_f(t,u)\,du = \sigma_r(T-t)$$

- The constancy of $\sigma(t,T)$ is unrealistic and provides too few degrees of freedom for calibration to quoted option prices.
- Interest rates are Gaussian, so they can become negative.
- With only one driving Brownian motion, the instantaneous moves of all forward rates are perfectly correlated (this is obviously common to all one-factor models.

4.3 The mean-reverting GSR model

Empirical studies suggest that interest rates exhibit mean-reversion. For instance, Vasicek assumed that interest rates are governed by an SDE

$$dr(t) = \kappa(\theta - r(t))dt + \sigma_r dW(t)$$

Integrating the linear SDE it follows that

$$r(t) = \theta + (r(0) - \theta)e^{-\kappa t} + \sigma_r \int_0^t e^{-\kappa(t-s)\, dW(s)}$$

so $r(t)$ is a Gaussian random variable with moments

$$\begin{aligned} E[r(t)] &= \theta + [r(0) - \theta]e^{-\kappa t} \\ Var[r(t)] &= \frac{\sigma_r^2}{2\kappa}(1 - e^{-2\kappa t}) \end{aligned}$$

Allowing all constants to be time-dependent in the Vasicek model we obtain the **general short rate (GSR) model**

$$dr(t) = \kappa(t)\,[\theta(t) - r(t)]\,dt + \sigma_r(t)dW(t)$$

Recall (c.f. Lemma 4.5.4 in [AP10-1]) that within an HJM setting we have

$$\begin{aligned} df(t,T) &= \sigma_f(t,T)^t \int_t^T \sigma_f(t,u)\,du dt + \sigma_f(t,T)^t dW(t) \\ \sigma_f(t,T) &= \sigma_r(t)\exp\left(-\int_t^T \kappa(u)\,du\right) \end{aligned}$$

and in order to match the initial yield curve, we must have

$$\theta(t) = \frac{1}{\kappa(t)}\frac{\partial f(0,t)}{\partial t} + f(0,t) + \frac{1}{\kappa(t)}\int_0^t e^{-2\int_u^t \kappa(s)\,ds}\sigma_r^2(u)\,du \tag{30}$$

The term $\frac{\partial f(0,t)}{\partial t}$ can be a nuisance in applications where the initial forward curve is not smooth. In order to get rid of it we can perform a change of variables.

Lemma 4.3. *Defining*

$$x(t) := r(t) - f(0,t)$$

the GSR model becomes

$$\begin{aligned} dx(t) &= [y(t) - \kappa(t)x(t)]\,dt + \sigma_r(t)dW(t), \\ x(0) &= 0, \\ y(t) &= \int_0^t e^{-2\int_u^t \kappa(s)\,ds}\sigma_r^2(u)\,du \end{aligned}$$

The bond reconstitution formula is

$$\begin{aligned} P(t,T) &= \frac{P(0,T)}{P(0,t)}\exp\left(-x(t)G(t,T) - \frac{1}{2}y(t)G^2(t,T)\right) \qquad (31) \\ G(t,T) &= \int_t^T e^{-\int_t^u \kappa(s)\,ds}\,du \end{aligned}$$

Remark 4.1. Note that the term $G(t,T)$ can be expressed as

$$G(t,T) = [G(0,T) - G(0,t)]\,e^{\int_0^t \kappa(s)\,ds}$$

which is useful in grid-based numerical work.

4.3.1 Swaption pricing

Assume that the function $\kappa(t)$ is fixed exogenously (based on empirical observations). By (30), the function $\theta(t)$ is then uniquely fixed by the initial forward curve. In order to have a full model specification, it thus suffices to determine the function $\sigma_r(t)$, which in practice is done by calibration of the model to observed prices of liquid European options (i.e. caps and swaptions). In any Gaussian HJM model, caplets can be priced by simple Black-Scholes formulas, so:

- In this subsection we focus on pricing European swaptions in two different ways.
- In the next subsection we use these formulas to calibrate the model (i.e. to find $\sigma_r(t)$).

4.3.1.1 Jamshidian decomposition

For a payer swaption expiring at time T_0 with the underlying swap paying an annualized coupon c at times $T_1 < \cdots < T_N$, the swaption payout at time T_0 is

$$V_{swaption}(T_0) = \left(1 - P(T_0, T_N) - c\sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1})\right)^+, \qquad \tau_i = T_{i+1} - T_i$$

Jamshidian re-writes the swaption payout from an option on a sum of discount bonds to a sum of options on discount bonds.

First write $P(T_0, T_N) = P(T_0, T_N; x(T_0))$ and let x^* be the critical value for which the swap at time T_0 is exactly zero. This can be found by numerical root search on the equation

$$P(T_0, T_N, x^*) + c\sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1}, x^*) = 1$$

Also define the strikes

$$K_i = P(T_0, T_i, x^*)$$

From the expression of $P(t,T)$ in terms of $x(t)$ it follows that $P(T_0, T_i, x(T_0))$ is monotonically decreasing in $x(T_0)$, the swaption only pays out a positive amount if $x(T_0) > x^*$. After a few computations one concludes

$$\begin{aligned} V_{swaption}(T_0) &= \left(1 - P(T_0, T_N, x(T_0)) - c\sum_{i=0}^{N-1} \tau_i P(T_0, T_{i+1}, x(T_0))\right) 1_{\{x(T_0)>x^*\}} \\ &= (K_N - P(T_0, T_N, x(T_0)))^+ + c\sum_{i=0}^{N-1} \tau_i \left((K_{i+1} - P(T_0, T_{i+1}, x(T_0))\right)^+ \end{aligned} \tag{32}$$

so the swaption payout has now been decomposed into $N+1$ put options on ZCB's, which allows us to price the swaption in closed form.

- The use of root-search algorithms and the need to price a potentially large amount of ZCB's can be cumbersome.
- **Alternative**: Examine forward swap rate in appropriate annuity measure, introducing approximations as needed to make the dynamics tractable (see next subsection).

4.3.1.2 Swap rate approximation

Consider again a payer swaption expiring at time T_0 with the underlying swap paying an annualized coupon c at times $T_1 < \cdots < T_N$. Re-write the swaption payout as

$$V_{swaption}(T_0) = A(T_0)(S(T_0) - c)^+$$

where, recall

$$A(t) \stackrel{not}{=} A_{0,N}(t) = \sum_{i=0}^{N-1} \tau_i P(t, T_{i+1}), \quad S(t) \stackrel{not}{=} S_{0,N}(t) = \frac{P(t,T_0) - P(t,T_N)}{A(t)}$$

In the Q^A measure we have

$$V_{swaption}(0) = A(0)E^A\left[(S(T_0) - c)^+\right]$$

and it now suffices to establish tractable dynamics for $S(t)$ in Q^A. $S(t)$ is obviously a Q^A-martingale, and from the bond reconstitution formula (31)

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

it is clear that $S(t)$ is a deterministic function $S(t) = S(t, x(t))$, so from It's lemma

$$dS(t) = q(t, x(t))\sigma_r(t)dW^A(t), \qquad q(t,x) = \frac{\partial}{\partial x} \frac{P(t,T_0,x) - P(t,T_N,x)}{\sum_{i=0}^{N-1} \tau_i P(t, T_{i+1}, x)}$$

where we used the bond reconstitution formula (31) to express the $P(t, T_i)$'s as functions of x. Evaluating the partial derivatives one easily sees that

$$q(t,x) = -\frac{P(t,T_0,x)G(t,T_0) - P(t,T_N,x)G(t,T_N)}{A(t,x)} + \frac{S(t,x)}{A(t,x)} \sum_{i=0}^{N-1} \tau_i P(t,T_{i+1},x)G(t,T_{i+1})$$

The function $q(t, x(t))$ can be experimentally verified to be close to a constant in the x-direction, so

$$q(t, x(t)) \simeq q(t, \bar{x}(t)), \quad \bar{x}(t) \text{ deterministic proxy for } x(t)$$

For instance, setting $\bar{x}(t) \equiv 0$ will yield good precision if $\sigma_r(t)$ is not too high. More accurate approximation will be given later on.

We are thus left with the problem

$$\begin{aligned} V_{swaption}(0) &= A(0)E^A\left[(S(T_0) - c)^+\right], \\ dS(t) &= q(t, \bar{x}(t))\sigma_r(t)dW^A(t) \end{aligned}$$

The dynamics of $S(t)$ can be thought of as those of a CEV process

$$dS(t) = \lambda(t)S^p(t)dW^A(t), \quad \lambda(t) = q(t, \bar{x}(t))\sigma_r(t)$$

with $p = 0$ and time-dependent volatility.

- By [AP10-1, Proposition 7.6.1], the prices of European options in the model $dS(t) = \lambda(t)S^p(t)dW^A(t)$ are the same as those of the time-averaged constant volatility model

$$dS(t) = \sqrt{\bar{\lambda}} S^p(t)dW^A(t), \quad \bar{\lambda} = \frac{1}{T_0}\int_0^{T_0} \lambda^2(u)\, du$$

- For $p = 0$ and constant λ, namely $dS(t) = \lambda dW^A(t)$, a standard computation (c.f. [AP10-1, Remark 7.2.9]) shows that

$$c(0, S(0); T_0, c, \lambda) = (S(0) - c)\mathcal{N}(d) + \lambda\sqrt{T_0}\mathfrak{n}(d), \quad d = \frac{S(0) - c}{\lambda\sqrt{T_0}}$$

In conclusion (simplifying $\sqrt{T_0}$) we obtain:

$$V_{swaption}(0) = A(0)\left[(S(0) - c)\mathcal{N}(d) + \sqrt{\bar{\lambda}} \cdot \mathfrak{n}(d)\right] \tag{33}$$

where

$$d = \frac{S(0) - c}{\sqrt{\bar{\lambda}}}, \quad \bar{\lambda} = \int_0^{T_0} q(t, \bar{x}(t))^2 \sigma_r(t)^2 \, dt$$

4.3.2 Swaption calibration

Using the option pricing formulas from the previous subsection, we can calibrate the model (i.e. find $\sigma_r(t)$ to match the market prices of one or more calibration targets, generally European swaptions).

Assume that we are given a collection[6] of $N - 1$ swaptions defined on a maturity grid $0 = T_0 < T_1 < \cdots < T_N$ such that the i-th swaption expires at time T_i, for $i = 1, \ldots, N - 1$.

- Assume that the mean reversion speed $\kappa(t)$ is a constant and that it is fixed exogenously.
- Observe that the value of the swaption expiring at T_i depends on the volatility curve $\sigma_r(s)$ for $s \in [0, T_i]$ only[7].
- Assume that the volatility function $\sigma_r(t)$ is piecewise-constant on the maturity grid

$$\sigma_r(t) = \sum_{i=0}^{N-2} \sigma_i \mathbb{I}_{[T_i, T_{i+1}]}$$

Calibration can hence be performed **one swaption at a time** via a series of one-dimensional root searches as follows:

1. Assume that $\sigma_0, \ldots, \sigma_{i-1}$ have been computed.
2. Set the value σ_i such that the model price for the swaption expiring at T_{i+1} equals its market price, by numerically inverting equation (33)

$$V_{swaption}(0) = A(0)\left[(S(0) - c)\mathcal{N}(d) + \sqrt{\bar{\lambda}} \cdot \mathfrak{n}(d)\right]$$

[6] Such a collection is called a *swaption strip*.

[7] Indeed, the i-th swaption payout at time T_i

$$V_{swaption}(T_i) = \left(1 - P(T_i, T_N) - c \sum_{n=i}^{N-1} \tau_n P(T_i, T_{n+1})\right)^+$$

so our claim follows from the bond reconstitution formula (31)

$$\begin{aligned} P(T_i, T) &= \frac{P(0,T)}{P(0,T_i)} \exp\left(-x(T_i)G(T_i, T) - \frac{1}{2} y(T_i) G^2(T_i, T)\right) \\ G(t, T) &= \int_t^T e^{-2\int_u^t \kappa(s)\, ds} \, du \\ y(t) &= \int_0^t e^{-\int_u^t \kappa(s)\, ds} \sigma_r^2(u) \, du \end{aligned}$$

3. Repeat Step 2 for $i = 0, \dots, N-2$.

More concretely:

(i) Pre-compute the terms

$$G(T_i, T_j) = \int_{T_i}^{T_j} e^{-\int_{T_i}^{u} \kappa(s)\, ds}\, du, \quad i \leq j$$

via the recursion

$$G(t,T) = [G(0,T) - G(0,t)]\, e^{\int_0^t \kappa(s)\, ds}$$

At the n-th step of the recursion, for $n = 1, \dots, N-1$, we seek to obtain σ_{n-1} so that the n-th swaption market price $\hat{V}_n$ is matched.

(ii) Evaluate $y(T_n)$. This is done via the recursion

$$\begin{aligned} y(T_n) &= e^{-2\kappa_{n-1}(T_{n-1}-T_{n-2})} y(T_{n-1}) + \frac{1}{2}\sigma_{n-1}^2 (T_{n-1} - T_{n-2}) \left[e^{-2\kappa_{n-1}(T_{n-1}-T_{n-2})} + 1 \right], \\ y(T_0) &= 0 \end{aligned}$$

which is obtained by applying the trapezoidal quadrature rule to the integral

$$y(t) = \int_0^t e^{-2\int_u^t \kappa(s)\, ds} \sigma_r(u)\, du$$

Note that σ_{n-1} is unknown at the n-th step: we are optimizing for it.

(iii) Use $y(T_n)$ to calculate the discount factors $P(T_n, T_j)$, for $j > n$ using

$$P(T_n, T_j) = \frac{P(0,T_j)}{P(0,T_n)} e^{-\frac{1}{2} y(T_n) G(T_n, T_j)^2}$$

(iv) Compute $A_n(t) \equiv A_{n,N}(t)$ and $S_n(t) \equiv S_{n,N}(t)$ at $t = T_0, \dots, T_n$.

(v) Compute $q(t)$ for $t = T_0, \dots, T_n$.

(vi) Compute $\bar{\lambda}^2 = \frac{1}{T_n} \int_0^{T_n} q(t, \bar{x}(t))^2 \sigma_r(t)^2\, dt$ using the trapezoidal rule (we are using the approximation $\bar{x}(t) \equiv 0$

$$\bar{\lambda} \simeq \sigma_0^2 (T_1 - T_0) \frac{q(T_0,0)^2 + q(T_1,0)^2}{2} + \cdots \sigma_{n-1}^2 (T_{n-1} - T_n) \frac{q(T_{n-1},0)^2 + q(T_n,0)^2}{2}$$

(vii) Use a one-dimensional root-search procedure to find σ_{n-1} satisfying equation (33)

$$\hat{V}_n(0) = A_n(0) \left[(S_n(0) - c) \mathcal{N}(d(\sigma_{n-1})) + \sqrt{\bar{\lambda}(\sigma_{n-1})} \cdot \mathfrak{n}(d(\sigma_{n-1})) \right]$$

4.3.3 Monte Carlo simulation

To price a derivative security paying $V(T)$ at time T, recalling that $x(t) = r(t) - f(0,t)$, we need to compute

$$V(0) = E^Q \left[V(T) e^{-\int_0^T r(u)\, du} \right] = P(0,T) E^Q \left[V(T, \{x(t)\}_{t=0}^T) e^{-\int_0^T x(u)\, du} \right] \tag{34}$$

The dynamics of $x(t)$ in the GSR model are

$$\begin{aligned} dx(t) &= [y(t) - \kappa(t) x(t)] dt + \sigma_r(t) dW(t), \\ y(t) &= \int_0^t e^{-2\int_u^t \kappa(s)\, ds} \sigma_r(u)^2\, du \end{aligned}$$

Discretizing into a schedule[8] $t_0 < t_1 < \cdots < t_N$ we see that

$$\begin{aligned}
x(t_{i+1}) &= E[x(t_{i+1})|x(t_i)] + \sqrt{\mathrm{Var}(x(t_{i+1})|x(t_i))}Z_i, \quad i = 0, \ldots, N-1, \\
E[x(t_{i+1})|x(t_i)] &= e^{-\int_{t_i}^{t_{i+1}} \kappa(u)\,du} x(t_i) + \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \kappa(u)\,du} y(s)\,ds, \\
\mathrm{Var}(x(t_{i+1})|x(t_i)) &= \int_{t_i}^{t_{i+1}} \left(e^{-\int_s^{t_{i+1}} \kappa(u)\,du} \sigma_r(s)\right)^2 ds
\end{aligned}$$

where $Z_0, \ldots, Z_{N-1}$ is a sequence of independent standard Gaussian random variables.

For each simulated path $x(t_0), \ldots, x(t_N)$, we can compute $V(T, \{x(t)\}_{t=0}^T)$. As for the quantity $I(T) = -\int_0^T x(u)\,du$ that is also present in (34), we can proceed via a standard quadrature (which will introduce a discretization bias) or we can use [AP10-2, Lemma 10.1.11] to simulate it as

$$I(t_{i+1}) = E[I(t_{i+1})|I(t_i), x(t_i)] + \sqrt{\mathrm{Var}(I(t_{i+1})|I(t_i), x(t_i))}\tilde{Z}_i, \quad i = 0, \ldots, N-1$$

where

$$\begin{aligned}
E[I(t_{i+1})|I(t_i), x(t_i)] &= I(t_i) - x(t_i)G(t_i, t_{i+1}) - \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} e^{-\int_s^u \kappa(v)\,dv} y(s)\,ds\,du, \\
\mathrm{Var}(I(t_{i+1})|I(t_i), x(t_i)) &= 2\int_{t_i}^{t_{i+1}} \int_{t_i}^{u} e^{-\int_s^u \kappa(v)\,dv} y(s)\,ds\,du - y(t_i)G(t_i, t_{i+1})^2, \\
\mathrm{Cov}(x(t_{i+1}), I(t_{i+1})|I(t_i), x(t_i)) &= -\int_{t_i}^{t_{i+1}} \int_{t_i}^{u} \sigma_r(s)^2 e^{-\int_s^u \kappa(v)\,dv} e^{-\int_s^{t_{i+1}} \kappa(v)\,dv}\,ds\,du
\end{aligned}$$

where $\tilde{Z}_0, \ldots, \tilde{Z}_{N-1}$ are independent Gaussian variables satisfying

$$\rho_i \overset{not}{=} \mathrm{Corr}(Z_i, \tilde{Z}_i) = \frac{\mathrm{Cov}(x(t_{i+1}), I(t_{i+1})|I(t_i), x(t_i))}{\sqrt{\mathrm{Var}(I(t_{i+1})|I(t_i), x(t_i))}\sqrt{\mathrm{Var}(x(t_{i+1})|I(t_i), x(t_i))}}$$

Using the Cholesky decomposition as usual, the sequence $\tilde{Z}_i$ can be constructed by generating, for each i, two independent Gaussian variables Z_i, Z_i^* and setting

$$\tilde{Z}_i = \rho_i Z_i + \sqrt{1 - \rho_i^2} Z_i^*$$

4.4 The affine one-factor model

The GSR model suffers from two main deficiencies:

(i) Interest rates distribute as normal variables, so they can become negative.

(ii) The short rate volatility $\sigma_r(t)$ is independent of the short rate itself, so there is no way to control the volatility skew implied by the model.

Both issues are addressed by the affine short rate model

$$dr(t) = \kappa(t)\left[\theta(t) - r(t)\right] dt + \sigma\sqrt{\alpha + \beta r(t)}dW(t)$$

[8] In order to see this, simply write down the solution to the linear SDE for $x(t)$. The choice of the schedule $t_0 < t_1 < \cdots < t_N$ depends on the particulars of the payoff. For instance, if $V(T)$ only depends on the yield curve at time T, we may take $N = 1$.

4.4.1 Bond pricing

In order to compute bond prices

$$P(t,T) = E_t\left[e^{-\int_t^T r(u)\,du}\right]$$

one establishes more generally the extended transform

$$g(t,T;c_1,c_2) = E_t\left[e^{-c_1 r(T) - c_2\int_t^T r(u)\,du}\right], \qquad c_1,c_2 \in \mathbb{C}$$

Lemma 4.4. *Wherever the extended transform $g(t,T;c_1,c_2)$ us defined, it is given by*

$$g(t,T;c_1,c_2) = \exp\left[A(t,T;c_1,c_2) - B(t,T;c_1,c_2)r(t)\right]$$

where A,B satisfy a system of Riccati ODE's

$$\begin{aligned} \frac{dA}{dt} - \kappa(t)\theta(t)B + \frac{1}{2}\sigma^2(t)\alpha B^2 &= 0, \\ -\frac{dB}{dt} + \kappa(t)B + \frac{1}{2}\sigma^2(t)\beta B^2 &= c_2 \end{aligned}$$

subject to the terminal conditions $A(T,T;c_1,c_2) = B(T,T;c_1,c_2) = 0$.

- This system of ODE's can be solved robustly by Runge-Kutta methods.
- If the parameters are piecewise constant, A,B can also be found analytically (c.f. section 10.2.2.2 in [AP10-2]).

4.4.2 Discount bond calibration

In the affine SDE

$$dr(t) = \kappa(t)\left[\theta(t) - r(t)\right]dt + \sigma\sqrt{\alpha + \beta r(t)}dW(t)$$

- The role of $\theta(t)$ is to calibrate the model to the initial term structure of discount bonds.
- The short rate volatility $\sigma(t)$ will be determined through calibration against swaptions and caps/floors.

As in the GSR model, it is convenient to eliminate the dependence of θ on $\frac{\partial f(0,t)}{\partial t}$. This is accomplished via the change of variables $x(t) = r(t) - f(0,t)$; the SDE for $x(t)$ is

$$\begin{aligned} dx(t) &= dr(t) - \partial_t f(0,t)dt = \left(w(t) - \kappa(t)x(t)\right)dt + \sigma(t)\sqrt{\xi(t) + \beta x(t)}dW(t) \qquad (35) \\ w(t) &= \kappa(t)\theta(t) - \partial_t f(0,t) - \kappa(t)f(0,t) \\ \xi(t) &= \alpha + \beta f(0,t) \end{aligned}$$

The extended transform $g(t,T;c_1,c_2)$ is then expressed in terms of $x(t)$

$$g(t,T;c_1,c_2) = e^{-c_1 f(0,T)}\frac{P(0,T)^{c_2}}{P(0,t)^{c_2}}\exp\left[C(t,T;c_1,c_2) - x(t)B(t,T;c_1,c_2)\right]$$

where B,C satisfy a system of Riccati ODE's

$$\begin{aligned} \frac{dC}{dt} - w(t)B + \frac{1}{2}\sigma^2(t)\xi(t)B^2 &= 0, \quad \xi(t) = \alpha + \beta f(0,t), \\ -\frac{dB}{dt} + \kappa(t)B + \frac{1}{2}\sigma^2(t)\beta B^2 &= c_2 \end{aligned}$$

4.4.2.1 Algorithm for $\omega(t)$, for fixed α, β and known $\kappa(t), \sigma(t)$

The function $\omega(t)$ is determined by matching observed discount bond prices at time 0. Concretely, $\omega(t)$ is specified on a grid $t_0 < \cdots < T_N$ under a piecewise-constant assumption $\omega(t) = \sum_{i=0}^{N-1} \omega_i \mathbb{I}_{(t_i, t_{i+1}]}(t)$.

Note that

$$P(t,T) = g(t,T;0,1) = \frac{P(0,T)}{P(0,t)} e^{c(t,T) - x(t)b(t,T)} \tag{36}$$

where

$$\begin{aligned} \frac{dc(t,T)}{dt} - \omega(t)b(t,T) + \frac{1}{2}\sigma^2(t)\xi(t)b^2(t,T) &= 0, && (37) \\ -\frac{db(t,T)}{dt} + \kappa(t)b(t,T) + \frac{1}{2}\sigma^2(t)\beta b^2(t,T) &= 1 && (38) \end{aligned}$$

subject to $c(T,T) = b(T,T) = 0$.

Setting $t = 0$ in equation (36) establishes the fundamental calibration requirement

$$c(0,T;\omega(\cdot)) = 0, \quad \forall T$$

The idea is then to solve equation (38) for $b(t_i, t_j)$, $j < i$ and then find w_i iteratively from equation (37) by integrating it over $[0, t_{i+1}]$, and using the fact that $\int_0^{t_{i+1}} \frac{dc(t,t_{i+1})}{dt} dt = -c(0, t_{i+1}) = 0$. Concretely:

1. As a pre-processing step, find $b(t_i, t_j)$ for all $j < i$ by solving equation (38).
2. For a given i, assume that ω_j is known for $j < i$, namely that $\omega(t)$ has been specified over $[0, t_i]$.
3. Integrating equation (37) over $[0, t_{i+1}]$ yields

$$\begin{aligned} -c(0, t_{i+1}) &= 0 \\ &= \overbrace{-\frac{1}{2}\int_0^{t_{i+1}} \sigma^2(s)\xi(s)b^2(s,t_{i+1})\, ds + \int_0^{t_i} \omega(s)b(s,t_{i+1})\, ds}^{\stackrel{not}{=}\Theta(t_i)} + \int_{t_i}^{t_{i+1}} \omega(s)b(s,t_{i+1})\, ds \\ &= \Theta(t_i) + \omega_i \int_{t_i}^{t_{i+1}} b(s,t_{i+1})\, ds \end{aligned}$$

   whereby one can solve for ω_i and then repeat the process for $i = 0, 1, \ldots, N-1$.

4.4.3 European option pricing

In order to to determine the volatility function $\sigma(t)$), the affine model is calibrated to caplets and swaptions, so we need efficient schemes to compute prices thereof.

4.4.3.1 Caplets

Recall that, by Lemma 4.4, the moment generating function is available in the affine model

$$g(t,T;c_1,c_2) = \exp\left[A(t,T;c_1,c_2) - B(t,T;c_1,c_2)r(t)\right]$$

where A, B satisfy a system of Riccati ODE's, so caplets (i.e. options on zero-coupon bonds, can be priced by Fourier methods.

Briefly: given a tenor structure $0 \leq T_0 < \cdots < T_N$, a caplet fixing at T_n has time-T_{n+1} payout

$$V_{caplet}(T_{n+1}) = \tau_n \left(L_n(T_n) - k\right)^+ = \left(\frac{1}{P(T_n, T_{n+1})} - 1 - \tau k\right)^+$$

So it's time-T_n value is

$$V_{caplet}(T_n) = P(T_n, T_{n+1}) E_{T_n}^{T_{n+1}} \left[\left(\frac{1}{P(T_n, T_{n+1})} - 1 - \tau k\right)^+\right] \overset{[*]}{=} \left(1 - (1 + k\tau)P(T_n, T_{n+1})\right)^+$$

where in $[*]$ we used that the expression inside the expectation is $\mathcal{F}_{T_n}$-measurable, whence the time-T_n value of the caplet can be represented as the payoff of a scaled put option on a zero-coupon bond for which we have an m.g.f. available.

4.4.3.2 Swaptions

In order to price swaptions efficiently, we work out an approximation for the swap rate martingale dynamics. Start by writing the swaption payout as

$$V_{swaption}(T_0) = A(T_0)(S(T_0) - c)^+$$

with

$$A(t) \overset{not}{=} A_{0,N}(t) = \sum_{i=0}^{N-1} \tau_i P(t, T_{i+1}), \quad S(t) \overset{not}{=} S_{0,N}(t) = \frac{P(t, T_0) - P(t, T_N)}{A(t)}$$

In the Q^A measure we have

$$V_{swaption}(0) = A(0) E^A \left[(S(T_0) - c)^+\right]$$

$S(t)$ is obviously a Q^A-martingale, and from the bond reconstitution formula (36)

$$P(t, T) = g(t, T; 0, 1) = \frac{P(0, T)}{P(0, t)} e^{c(t,T) - x(t)b(t,T)}$$

we obtain

$$dS(t) = \frac{\partial S(t, x(t))}{\partial x} \sigma(t) \sqrt{\xi(t) + \beta x(t)} dW^A(t)$$

where

$$\frac{\partial S(t, x)}{\partial x} = -\frac{b(t, T_0)P(t, T_0) - b(t, T_N)P(t, T_N)}{A(t)} + \frac{S(t)}{A(t)} \sum_{i=0}^{N-1} \tau_i b(t, T_{i+1}) P(t, T_{i+1})$$

$S(t)$ can be empirically verified to be well approximated by a linear function of $x(t)$, say $S(t) = \zeta(t) + \chi(t)x(t)$, whence

$$dS(t) \simeq \sigma(t) \sqrt{\xi_s(t) + \beta_s(t) S(t)} dW^A(t), \quad \text{for appropriate } \xi_s, \beta_s(t)$$

We can further reduce to

$$dS(t) = \sigma(t) \sqrt{\beta_s(t)} \sqrt{\psi + S(t)} dW^A(t), \quad \psi = \frac{\int_0^{T_0} \sigma(t)^2 \beta_s(t) \xi_s(t)\, dt}{\int_0^{T_0} \sigma(t)^2 \beta_s(t)\, dt}$$

Defining $y(t) = \psi + S(t)$ our swaption pricing problem reduces to

$$\begin{aligned} V_{swaption}(0) &= A(0)E^A\left[(S(T_0)-c)^+\right] = A(0)E^A\left[(y(T_0)-(c+\psi))^+\right], \\ dy(t) &= \sigma(t)\sqrt{\beta_s(t)}\sqrt{y(t)}dW^A(t) \\ y(0) &= \psi + S(0) \end{aligned}$$

Now $y(t)$ is simply a CEV process with power $p = \frac{1}{2}$ and time dependent volatility $\lambda(t) := \sigma(t)\sqrt{\beta_s(t)}$. By [AP10-1, Proposition 7.6.1], the swaption price will be that of a swaption in a CEV with constant time-averaged volatility

$$dy(t) = \sqrt{\bar{\lambda}}\sqrt{y(t)}dW^A(t), \quad \bar{\lambda} = \frac{1}{T_0}\int_0^{T_0} \sigma(u)\sqrt{\beta_s(u)}\,du$$

By [AP10-1, Proposition 7.2.6], the swaption price will be

$$V_{swaption}(0) = V_{swaption}(0, y(0); T_0, \psi + c) = y(0)\left[1 - \Upsilon(a, b+2, c)\right] - (\psi + c)\Upsilon(c, b, a)$$

where

$$a = \frac{4(\psi + c)}{\bar{\lambda}^2 T_0}, \quad b = 2, \quad c = \frac{4y(0)}{\bar{\lambda}^2 T_0}$$

where $\Upsilon(x, \nu, \gamma) = P(\chi^2_\nu(\gamma) \le x)$ is the cumulative distribution function for a non-central chi-square distribution $\chi^2_\nu(\gamma)$ with ν degrees of freedom and non-centrality parameter γ.

4.4.4 Swaption calibration

We finally illustrate how to calibrate the volatility function $\sigma(t)$ to a swaption strip defined on a maturity grid $0 = T_0 < \cdots < T_N$. Assume that all swaptions are written on swaps that mature at time T_N...

4.5 Log-normal short-rate models

Some authors have attempted to specify short-rate models where the dynamics of $r(t)$ are of the form

$$dr(t) = \mathcal{O}(dt) + \sigma_r(t)r(t)dW(t)$$

A few approaches are outlined in [AP10-2, Section 11.1]:

- Black-Derman-Toy model

  $$r(t) = U(t)e^{\sigma_r(t)W(t)}, \quad U(t), \sigma_r(t) \text{ deterministic}$$

  This model lacks a bond reconstitution formula and additionally, $\ln r(t)$ can be seen to be a mean-reverting Gaussian [AP10-2, Lemma 11.1.1] with a mean reversion speed $\kappa(t) = -\frac{\sigma_r'(t)}{\sigma_r(t)}$ that is beyond user control.
- The Black-Karasinski model is analogous to the BDT model, but it specified the mean reersion speed exogenously.

A common problem shared by all these log-normal models is that

$$E_t\left[\frac{1}{P(t', T)}\right] = \infty, \quad t < t' < T$$

In particular since $L(t, T) = \frac{1}{\tau}\left(\frac{P(t,T)}{P(t,T+\tau)} - 1\right)$, this implies the following unfortunate economic corollaries:

- $E_t[L(T,T)] = \infty$ for $T > t$, which predicts that all future Libor rates should be infinite.
- Using Jensen's inequality we obtain

$$\frac{1}{P(t',T)} = \frac{1}{E_{t'}\left[e^{-\int_{t'}^{T} r(u)\,du}\right]} \leq E_{t'}\left[e^{\int_{t'}^{T} r(u)\,du}\right]$$

  whence $E_t\left[e^{\int_{t'}^{T} r(u)\,du}\right] = \infty$, so that the expected return of investing in the continuously compounded money market account for a finite period of time is infinite.

4.6 Spanned and unspanned stochastic volatility

We illustrate why the short rate framework is not particularly amenable to stochastic volatility extensions. Consider the Fong-Vasicek model

$$\begin{aligned} dr(t) &= \kappa_r(\theta_r - r(t))dt + \sqrt{z(t)}dW_1(t), \quad \kappa_r, \theta_r \in \mathbb{R}^+ \\ dz(t) &= \kappa_z(\theta_z - z(t))dt + \eta\sqrt{z(t)}dW_2(t), \quad \kappa_z, \theta_z \in \mathbb{R}^+ \\ dW_1(t) \cdot dW_2(t) &= \rho dt \end{aligned}$$

The bond prices in this model can be seen to satisfy

$$P(t, t+\delta) = e^{A(\delta)+r(t)B(\delta)+z(t)C(\delta)}$$

where A, B, C a deterministic functions satisfying a coupled system of ODE's. Therefore, $P(t,T)$ is a *deterministic function of the two state variables* $r(t)$ and $z(t)$ and hence the exposure to both can be hedged out by taking positions in two discount bonds with different maturities[9]. When this happens, we say that the state variables are *spanned by the discount curve.*

There is evidence, however, that interest rate option volatilities cannot be perfectly hedged by trading only discount bonds, which implies that volatilities of discount bonds depend on a vector of random state variables $(z_1(t), \ldots, z_n(t))$ that are not included in the state variables used in reconstitution formulas for the discount curve, namely

$$\frac{\partial P(t,T)}{\partial z_i(t)} = 0, \quad i = 1, \ldots, n$$

[9]Namely, given observations at time t of the prices of two discount bonds with different maturities, we can invert the reconstitution formula for $P(t,T)$ to uncover the current values of the two state variables.

5 Multi-factor short-rate models

Short rate models with a single driving Brownian motion:

1. Imply that the instantaneous correlation between forward rates at different maturities is one, which contradicts reality. As a general rule, all derivatives that have payouts exhibiting significant convexity to non-parallel moves of the forward curve must *not* be priced in a one-factor model.
2. Are unable to produce a time-stationary non-monotonic term structure of forward rate volatilities (the so-called short-maturity volatility hump). Two-factor Gaussian models are, in contrast, capable of reproducing this volatility hump.

In this section we briefly sketch:

- The Gaussian multi-factor model, following [AP10-2, Section 12.1]: development from separability condition, swaption pricing and model calibration.
- The quasi-Gaussian multi-factor model, following [AP10-2, Section 13.3]

5.1 The Gaussian multi-factor model

One starts along the lines of the one-factor Gaussian model. A general d-factor model can be written as

$$\begin{aligned} df(t,T) &= \sigma_f(t,T)^\top \int_0^t \sigma_f(t,u)\,du dt + \sigma_f(t,T)^\top dW(t), \\ \frac{dP(t,T)}{P(t,T)} &= r(t)dt - \sigma_P(t,T)^\top dW(t) \end{aligned}$$

where $\sigma_P(t,T)$ is a bounded d-dimensional function of time and $W(t)$ is a d-dimensional Brownian motion in the risk-neutral measure Q. This model can be made Markovian by imposing a separability condition:

Proposition 5.1. *Assuming that*

$$\sigma_f(t,T) = g(t)h(T) \tag{39}$$

where g is a $d\times d$ deterministic matrix-valued function, and h is a d-dimensional deterministic vector. Then

$$\begin{aligned} f(t,T) &= f(0,T) + \Omega(t,T) + h(T)^\top z(t), &(40)\\ dz(t) &= g(t)^\top dW(t), \quad z(0) = 0, &(41)\\ r(t) &= f(0,t) + \Omega(t,T) + h(t)^\top z(t), &(42)\\ \Omega(t,T) &= h(T)^\top \int_0^t g(s)^\top g(s) \int_s^T h(u)\,du\,ds &(43) \end{aligned}$$

If the separability condition is satisfied, the *forward curve can be reconstructed from d Gaussian martingale variables $z_i(t)$*, which are not determined uniquely. Set

$$H(t) = \mathrm{diag}(h(t)) \equiv \mathrm{diag}(h_1(t), \ldots, h_d(t)), \quad h_i(t) \neq 0 \quad i = 1, \ldots, d$$

which is clearly invertible, and define a diagonal $d \times d$ matrix[10] by

$$\kappa(t) = -\frac{dH(t)}{dt} H(t)^{-1} \tag{44}$$

[10] Recall that in the HJM setting $\kappa(t) = -\frac{h'(t)}{h(t)}$, where $\sigma_f(t,T) = g(t)h(T)$.

Proposition 5.2. *Let the forward rate volatility by separable $\sigma_f(t,T) = g(t)h(T)$. Let $\kappa(t)$ be defined as above and define*

$$x(t) = H(t)\int_0^t g(s)^\top g(s)\int_s^t h(u)\,du\,ds + H(t)z(t) \tag{45}$$

$$y(t) = H(t)\left(\int_0^t g(s)^\top g(s)\,ds\right)H(t) \tag{46}$$

(so that $x(t)$ is a d-dimensional random vector and $y(t)$ is a deterministic $d\times d$ symmetric matrix). Then if $\mathbf{1} = (1,1,\dots,1)^\top \in \mathbb{R}^d$ we have

$$\begin{aligned} dx(t) &= [y(t)\mathbf{1} - \kappa(t)\cdot x(t)]\,dt + \sigma_x(t)^\top dW(t), \\ \sigma_x(t) &= g(t)H(t) \end{aligned}$$

and with $M(t,T) := H(T)H(t)^{-1}\mathbf{1}$,

$$f(t,T) = f(0,T) + M(t,T)^\top\left(x(t) + y(t)\int_t^T M(t,u)\,du\right) \tag{47}$$

$$r(t) = f(t,t) = f(0,t) + \mathbf{1}^\top x(t) = f(0,t) + \sum_{i=1}^d x_i(t) \tag{48}$$

As for the bond prices, defining

$$G(t,T) = \int_t^T M(t,u)\,du$$

we can write explicitly

$$P(t,T) = \frac{P(0,T)}{P(0,t)}\exp\left[-G(t,T)^\top\cdot x(t) - \frac{1}{2}G(t,T)^\top\cdot y(t)\cdot G(t,T)\right] \tag{49}$$

More explicitly: writing

$$\kappa(t) = \mathrm{diag}(\kappa_1(t),\dots,\kappa_d(t))$$

it follows from the definitions above that

$$h(t) = \left(e^{-\int_0^t \kappa_1(s)\,ds},\dots,e^{-\int_0^t \kappa_d(s)\,ds}\right)^\top \tag{50}$$

$$M(t,T) = H(T)H(t)^{-1}\mathbf{1} = \left(e^{-\int_t^T \kappa_1(s)\,ds},\dots,e^{-\int_t^T \kappa_d(s)\,ds}\right)^\top \tag{51}$$

$$y(t) = \int_0^t \mathrm{diag}(M(s,t))\cdot\sigma_x(s)^\top\cdot\sigma_x(s)\cdot\mathrm{diag}(M(s,t))\,ds \tag{52}$$

whereby specification of the mean reversion $\kappa(t)$ and $\sigma_x(t)$ fully determines the d-dimensional Gaussian model.

One important motivation for the introduction of a multi-factor interest rate model is the ability to control correlations among the various points on the forward curve.

Lemma 5.3. *In the model for $r(t)$ described in Proposition 5.1, the time-t instantaneous correlation between the forward rates $f(t,T_1)$ and $f(t,T_2)$ is given by*

$$\rho(t,T_1,T_2) = \frac{h(T_1)^\top\cdot g(t)^\top\cdot g(t)\cdot h(T_2)}{\sqrt{h(T_1)^\top\cdot g(t)^\top\cdot g(t)\cdot h(T_1)}\cdot\sqrt{h(T_2)^\top\cdot g(t)^\top\cdot g(t)\cdot h(T_2)}} \tag{53}$$

5.2 Two-factor Gaussian model

In this case we have

$$h(t) = \begin{pmatrix} e^{-\int_0^t \kappa_1(u)\,du} \\ e^{-\int_0^t \kappa_2(u)\,du} \end{pmatrix}, \quad g(t) = \begin{pmatrix} \sigma_{11}(t)e^{-\int_0^t \kappa_1(u)\,du} & 0 \\ \sigma_{21}(t)e^{-\int_0^t \kappa_1(u)\,du} & \sigma_{22}(t)e^{-\int_0^t \kappa_2(u)\,du} \end{pmatrix}$$

and writing $x(t) = (x_1(t), x_2(t))^\top$

$$dx(t) = [y(t)\mathbf{1} - \kappa(t)x(t)]dt + \sigma_x(t)^\top dW(t), \quad \sigma_x(t) = \begin{pmatrix} \sigma_{11}(t) & 0 \\ \sigma_{21}(t) & \sigma_{22}(t) \end{pmatrix}$$

The instantaneous correlation between $x_1(t)$ and $x_2(t)$ is

$$\rho_x(t) = \frac{\sigma_{22}(t)\sigma_{21}(t)}{\sqrt{\sigma_{11}(t)^2 + \sigma_{21}(t)^2} \cdot \|\sigma_{22}(t)\|}$$

It is clear that the model has the forward rate process

$$df(t,T) = \mathcal{O}(dt) + \begin{pmatrix} \sigma_1(t)e^{-\int_t^T \kappa_1(u)\,du} \\ \sigma_2(t)e^{-\int_t^T \kappa_2(u)\,du} \end{pmatrix} dW^*(t), \quad dW_1^*(t)dW_2^*(t) = \rho_x(t)dt$$

with $\sigma_1(t) = \sqrt{\sigma_{11}(t)^2 + \sigma_{21}(t)^2}$ and $\sigma_2(t) = \|\sigma_{22}(t)\|$. In particular (c.f. [AP10-2, Lemma 12.1.11]) we have

$$\rho(t, T_1, T_2) := \mathrm{Corr}_t(df(t,T_1), df(t,T_2)) = \frac{b(t,T_1,T_2)}{\sqrt{b(t,T_1,T_1)b(t,T_2,T_2)}} \tag{54}$$

where

$$b(t,T_1,T_2) = 1 + \rho_x(t)\frac{\sigma_2(t)}{\sigma_1(t)}\left(e^{-\int_t^{T_1}(\kappa_2(u)-\kappa_1(u))\,du} + e^{-\int_t^{T_2}(\kappa_2(u)-\kappa_1(u))\,du}\right)$$
$$+ \left(\frac{\sigma_2(t)}{\sigma_1(t)}\right)^2 e^{-\int_t^{T_1}(\kappa_2(u)-\kappa_1(u))\,du - \int_t^{T_2}(\kappa_2(u)-\kappa_1(u))\,du}$$

5.3 Swaption pricing (via Gaussian swap rate approximation)

Swaptions in a multi-factor Gaussian model can be priced identically to the one-dimensional setting (c.f. Section 4.3.1.2).

The time-0 value of a swaption is given by $V_{swaption}(0) = A(0)E^A\left[(S(T_0) - c)^+\right]$ The swap rate dynamics can be approximated as

$$dS(t) = \boldsymbol{q}(t, x(t))^\top \boldsymbol{\sigma_x}(t)^\top d\boldsymbol{W}^A(t), \quad q_j(t, \boldsymbol{x}(t)) = \frac{\partial S(t)}{\partial x_j}$$

The functions q_j can be experimentally verified to be close to a constant, so we can approximate

$$q_j(t, \boldsymbol{x}(t)) \simeq q_j(t, \bar{\boldsymbol{x}}(t))$$

where $\bar{\boldsymbol{x}}(t)$ is some deterministic proxy of the random vector $\boldsymbol{x}(t)$, for instance $\bar{\boldsymbol{x}}(t) = \mathbf{0}$. Then, as in the 1-dimensional case we have the following simple swaption pricing formula.

Lemma 5.4. *Let $\bar{\boldsymbol{x}}(t)$ be a deterministic function of time and assume that $q_j(t, \boldsymbol{x}(t)) \simeq q_j(t, \bar{\boldsymbol{x}}(t))$. Then*

$$V_{swaption}(0) = A(0)\left[(S(0) - c)\mathcal{N}(d) + \sqrt{\bar{\lambda}}\mathfrak{n}(d)\right]$$

where

$$d = \frac{S(0) - c}{\sqrt{\bar{\lambda}}}, \quad \bar{\lambda} = \int_0^{T_0} \|\boldsymbol{q}(t, \bar{x}(t))^\top \boldsymbol{\sigma_x}(t)^\top\|^2\,dt$$

5.3.1 Calibration

In order to calibrate the Gaussian multi-factor model we need to specify the functions $g(\cdot)$ ad $h(\cdot)$. From (50)

$$\boldsymbol{h}(t) = \left(e^{-\int_0^t \kappa_1(s)\,ds}, \ldots, e^{-\int_0^t \kappa_d(s)\,ds}\right)^\top$$

it is clear that specification of the mean reversions $(\kappa_1(t), \ldots, \kappa_d(t)))$ fully determines the function $h(\cdot)$.

The functions $g(\cdot)$ ad $h(\cdot)$ affect both volatilities and correlations of market rates, so in order to calibrate them to market instrument prices, we need both caplet/swaptions and also correlation-sensitive instruments like spread options. The former are far more liquid than the latter, so it is a good idea to try to separate volatility and correlation calibration.

Consider the following:

- d benchmark tenors $\delta_1 < \cdots < \delta_d$.
- d forward rates $f_i(t) = f(t, t+\delta_i)$, $i = 1, \ldots, d$, with instantaneous correlations[11]

$$\boldsymbol{\lambda}_i(t) = \boldsymbol{h}(t+\delta_i)^\top \boldsymbol{g}(t)^\top, \quad \lambda_i(t) := \|\boldsymbol{\lambda}_i(t)\|$$

- Denote by $\chi_{i,j}(t) = \mathrm{Corr}(f_i(t), f_j(t))$, $i, j = 1, \ldots, d$.
- Denote the covariance matrix of the vector $\boldsymbol{f}(t) = (f_1(t), \ldots, f_d(t))^\top$ by

$$[\boldsymbol{R}^{\boldsymbol{f}}(t)]_{i,j} = \boldsymbol{\lambda}_i(t)\boldsymbol{\lambda}_j^\top(t) = \chi_{i,j}(t)\lambda_i(t)\lambda_j(t)$$

  Also denote its Cholesky decomposition by

$$\boldsymbol{R}^{\boldsymbol{f}}(t) = \boldsymbol{C}^{\boldsymbol{f}}(t)^\top \boldsymbol{C}^{\boldsymbol{f}}(t)$$

- This instantaneous covariance matrix can also be written as

$$\boldsymbol{H}^{\boldsymbol{f}}(t)\boldsymbol{g}(t)^\top \boldsymbol{g}(t)\boldsymbol{H}^{\boldsymbol{f}}(t)^\top, \quad \boldsymbol{H}^{\boldsymbol{f}}(t) = \begin{pmatrix} \boldsymbol{h}(t+\delta_1)^\top \\ \vdots \\ \boldsymbol{h}(t+\delta_d)^\top \end{pmatrix} = \begin{pmatrix} h_1(t+\delta_1) & \cdots & h_d(t+\delta_1) \\ \vdots & \ddots & \vdots \\ h_1(t+\delta_d) & \cdots & h_d(t+\delta_d) \end{pmatrix}$$

  whence

$$\boldsymbol{H}^{\boldsymbol{f}}(t)\boldsymbol{g}(t)^\top = \boldsymbol{C}^{\boldsymbol{f}}(t) \Longrightarrow \boldsymbol{g}(t)^\top = \boldsymbol{H}^{\boldsymbol{f}}(t)^{-1}\boldsymbol{C}^{\boldsymbol{f}}(t)$$

- The matrix $\boldsymbol{g}(t)$ can also be determined by fitting the correlation rather than the covariance. Concretely, defining

$$[\boldsymbol{X}^{\boldsymbol{f}}(t)]_{i,j} := \chi_{i,j}(t) \quad \text{and} \quad \boldsymbol{X}^{\boldsymbol{f}}(t) = \boldsymbol{D}^{\boldsymbol{f}}(t)^\top \boldsymbol{D}^{\boldsymbol{f}}(t)$$

[11] Recall the d-factor Gaussian model specification from Proposition 5.1

$$\begin{aligned} f(t,T) &= f(0,T) + \Omega(t,T) + h(T)^\top z(t), \\ dz(t) &= g(t)^\top dW(t), \quad z(0) = 0, \\ \Omega(t,T) &= h(T)^\top \int_0^t g(s)^\top g(s) \int_s^T h(u)\,du\,ds \end{aligned}$$

whence for $T = t + \delta_i$ we have

$$df_i(t) = df_i(t, t+\delta_i) = d\Omega(t, t+\delta_i)dt + \boldsymbol{h}(t+\delta_i)^\top d\boldsymbol{z}(t) = d\Omega(t, t+\delta_i)dt + \underbrace{\boldsymbol{h}(t+\delta_i)^\top \boldsymbol{g}(t)^\top}_{:=\boldsymbol{\lambda}_i(t)} d\boldsymbol{W}(t).$$

one can obtain $\boldsymbol{g}(t)$ by solving

$$\boldsymbol{H^f}(t)\boldsymbol{g}(t)^\top = \mathrm{diag}(\lambda_1(t), \ldots, \lambda_d(t))\boldsymbol{D^f}(t)^\top$$

with the advantage that $\boldsymbol{D^f}(t)$ is independent of the volatilities $\lambda_i(t)$ and it doesn't need to be recomputed after each update of $\lambda_i(t)$, within a calibration iteration.

One can thus implement the following calibration algorithm:

1. Specify the vector $\boldsymbol{h}(t)$ by the mean-reversion parametrization (50)

$$\boldsymbol{h}(t) = \left(e^{-\int_0^t \kappa_1(s)\,ds}, \ldots, e^{-\int_0^t \kappa_d(s)\,ds}\right)^\top$$

   using d different constant mean reversions (expand on how to select κ_i.)
2. Fill in the correlation matrix $[\boldsymbol{X^f}(t)]_{i,j} = \chi_{i,j}(t)$, as explained later in Section 7.2.2.
3. Calibrate benchmark rate volatilities against swaptions using the pricing formula from Lemma 5.4.

   Concretely, consider the 2-factor case and assume that

$$\kappa(t) = (\kappa_1(t), \kappa_2(t))^\top, \quad \text{and} \quad R^f(t) = \begin{pmatrix} 1 & \chi_{12}(t) \\ \chi_{12}(t) & 1 \end{pmatrix}$$

   have been pre-specified. Consider two swaption strips (same fixing dates, different tenors) on a tenor structure $T_0 < T_1 < \cdots < T_N$. Assume that $\sigma_{11}(t), \sigma_{21}(t), \sigma_{22}(t)$ are piecewise constant, with $\sigma_{ij}(t) = \sum_{n=0}^{N-1} \sigma_{ij,n}\mathbb{I}_{[T_n,T_{n+1}]}$. For each $n = 1, \ldots, N-1$, use Lemma 5.4 to find the model price of the two swaptions maturing at T_n. These depend on three unknown variables $\sigma_{ij,n-1}(t)$, which are related by equation (54)

$$\mathrm{Corr}(df_1(t), df_2(t)) = \chi_{12}(t)$$

   This gives us three equations for three unknown at each step $n = 1, \ldots, N-1$.
4. Recover the diffusion matrix $\boldsymbol{g}(t)$ by solving

$$\boldsymbol{H^f}(t)\boldsymbol{g}(t)^\top = \mathrm{diag}(\lambda_1(t), \ldots, \lambda_d(t))\boldsymbol{D^f}(t)^\top$$

This procedure allows us to calibrate d swaption strips, ideally d strips with constant swap tenors matching the benchmark tenors $\delta_1, \ldots, \delta_d$.

5.3.2 Monte Carlo simulation

Monte Carlo methods for the d-dimensional Gaussian model are straightforward, as all state variables are jointly Gaussian. Recalling that $r(t) = f(0,t) + \mathbf{1}^\top x(t)$, for a security paying $V(T)$ at time T we need to compute

$$V(0) = E^Q\left[V(T)e^{-\int_0^T r(u)\,du}\right] = P(0,T)E^Q\left[V(T; \{x(t)\}_{t=0}^T)e^{-\int_0^T \mathbf{1}^\top x(u)\,du}\right]$$

The risk-neutral dynamics of $\boldsymbol{x}(t) = (x_1(t), \ldots, x_d(t))$ are

$$dx(t) = [y(t)\mathbf{1} - \kappa(t)x(t)]dt + \sigma_x(t)^\top dW(t), \quad x(0) = 0$$

Given a discrete schedule $\{t_i\}_{i=0}^N$, it is clear that $x(t_{i+1})|x(t_i)$ is d-dimensional Gaussian with mean and covariance matrix

$$\begin{aligned} E^Q\left[x(t_{i+1})|x(t_i)\right] &= e^{-\int_{t_i}^{t_{i+1}} \kappa(u)\,du}x(t_i) + \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \kappa(u)\,du}y(s)\mathbf{1}\,ds \\ \mathrm{Var}^Q\left[x(t_{i+1})|x(t_i)\right] &= \int_{t_i}^{t_{i+1}} e^{-\int_s^{t_{i+1}} \kappa(u)\,du}\sigma_x(s)\sigma_x(s)^\top e^{-\int_s^{t_{i+1}} \kappa(u)^\top\,du} \end{aligned}$$

If C is the square root of the covariance matrix, for a draw of d independent Gaussian samples $\boldsymbol{Z} = (Z_1, \ldots, Z_d)$ the advancement of $x(\cdot)$ from t_i to t_{i+1} is given by

$$x(t_{i+1}) = E^Q\left[x(t_{i+1})|x(t_i)\right] + CZ$$

As for the quantity $I(T) = -\int_0^T \mathbf{1}^\top x(u)\,du$:

- $I(t)$ is a Gaussian process, so we can simulate it jointly with $x(t)$ once we have worked out the moments of $I(t_{i+1})|I(t_i), x(t_i)$, as in section 4.3.3.
- We can compute $I(T)$ by numerical integration $I(T) \simeq -\mathbf{1}^\top \sum_{i=1}^N x(t_i)$.

6 The one-factor quasi-Gaussian model

We consider extensions of one-factor Gaussian short rate models with local and stochastic volatility, in which the Markovian structure of the model comes at the expense of adding additional state variables.

6.1 Definition

The Gaussian (GSR) model es given, within an HJM setting, by

$$
\begin{aligned}
dr(t) &= \kappa(t)\left[\theta(t) - r(t)\right] dt + \sigma_r(t) dW(t) \\
df(t,T) &= \sigma_f(t,T)^t \int_t^T \sigma_f(t,u)\, du dt + \sigma_f(t,T)^t dW(t) \\
\sigma_f(t,T) &= \sigma_r(t) \exp\left(-\int_t^T \kappa(u)\, du\right)
\end{aligned}
$$

Recall that in order to match the initial yield curve, we must have

$$
\theta(t) = \frac{1}{\kappa(t)} \frac{\partial f(0,t)}{\partial t} + f(0,t) + \frac{1}{\kappa(t)} \int_0^t e^{-2\int_u^t \kappa(s)\, ds} \sigma_r^2(u)\, du
$$

In order to make the short rate $r(t)$ Markovian we imposed a separability condition $\sigma_f(t,T) = g(t)h(T)$, where g, h are deterministic functions. Quasi-Gaussian models arise by allowing the function g to be stochastic:

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

Lemma 6.1 (description of quasi-Gaussian models)**.** *Consider a one-factor quasi-Gaussian model*

$$
\begin{aligned}
dr(t) &= \kappa(t)\left[\theta(t) - r(t)\right] dt + \sigma_r(t) dW(t) \\
df(t,T) &= \sigma_f(t,T)^t \int_t^T \sigma_f(t,u)\, du dt + \sigma_f(t,T)^t dW(t) \\
\sigma_f(t,T) &= \sigma_r(t) \exp\left(-\int_t^T \kappa(u)\, du\right) \\
dx(t) &= \left[y(t) - \kappa(t)x(t)\right] dt + \sigma_r(t,w) dW(t), \quad x(0) = 0 \\
dy(t) &= \left[\sigma_r^2(t,w) - 2\kappa(t)y(t)\right] dt, \quad y(0) = 0 \\
\kappa(t) &= -\frac{h'(t)}{h(t)} \\
\sigma_r(t,w) &= \sigma_f(t,t,w) = g(t,w)h(t)
\end{aligned}
$$

All zero-coupon discount bonds are deterministic functions of the processes $x(t), y(t)$

$$
P(t,T) = P(t,T,x(t),y(t))
$$

where

$$
\begin{aligned}
P(t,T,x,y) &= \frac{P(0,T)}{P(0,t)} \exp\left(-G(t,T)x - \frac{1}{2}G^2(t,T)y\right) \qquad (55) \\
G(t,T) &= \frac{1}{h(t)} \int_t^T h(s)\, ds
\end{aligned}
$$

Note that:

- The evolution of the whole interest rate curve can be reduced to the evolution of just two state variables $x(t)$ (main yield curve driver) and $y(t)$ (auxiliary variable required to uphold the no-arbitrage condition).
- The qG model has a closed-form bond reconstitution formula for arbitrary choices of $g(t,\omega)$ (we gain tractability at the expense of requiring two state variables, even though the Brownian motion is one-dimensional.

6.2 Local volatility

A one-factor gQ model with local volatility is obtained by requiring the function $g(\cdot)$ to be deterministic

$$g(t) = g(t, x(t), y(t)) \text{ and hence } \sigma_r(t) = g(t, x(t), y(t))h(t)$$

with dynamics

$$\begin{aligned} dx(t) &= [y(t) - \kappa(t)x(t)]\, dt + \sigma_r(t, x(t), y(t))dW(t), \\ dy(t) &= \left[\sigma_r^2(t, x(t), y(t)) - 2\kappa(t)y(t)\right] dt \end{aligned}$$

Fix a tenor structure $0 < T_0 < \cdots < T_N$, with $\tau_n = T_{n+1} - t_n$ and consider a forward swap rate[12] $S(t)$ with first fixing T_0 and last payment T_N. Since clearly $S(t) = S(t, x(t), y(t))$, using It's lemma and noting that $S(t)$ is a Q^A-martingale we conclude

$$dS(t) = \frac{\partial S}{\partial x}(t, x(t), y(t))\sigma_r(t, x(t), y(t))dW^A(t), \tag{56}$$

where

$$\begin{aligned} \frac{\partial S}{\partial x}(t, x, y) = -\frac{1}{A(t, x, y)} &\left[P(t, T_0, x, y)G(t, T_0) - P(t, T_N, x, y)G(t, T_N)\right] \\ &+ \frac{S(t, x, y)}{A(t, x, y)} \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}, x, y)G(t, T_{n+1}) \end{aligned} \tag{57}$$

Using the methods of *Markovian projection* one can show that the dynamics of the swap rate (56) can be written with a diffusion term that is just a function of the swap rate itself. The resulting SDE can hence be solved using standard methods for local volatility models.

Proposition 6.2 (Approximate local volatility dynamics for swap rate)**.** *The values of all European options on the swap rate $S(t)$ can be computed in a vanilla model with time-dependent local volatility function:*

$$dS(t) = \varphi(t, S(t))dW^A(t) \tag{58}$$

$$\varphi^2(t, s) = E^A\left[\left\{\frac{\partial S}{\partial x}(t, x(t), y(t))\sigma_r(t, x(t), y(t))\right\}^2 |S(t) = s\right] \tag{59}$$

Evaluating conditional expectations like these can be hard. However:

[12]Recall that, given a tenor structure $0 < T_0 < \cdots < T_N$, we have

$$A(t) \equiv A_{0,N}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}), \qquad S(t) \equiv S_{0,N}(t) = \frac{P(t, T_0) - P(t, T_n)}{A(t)}.$$

- If we assume that $y(t)$ is well approximated by a deterministic function $\bar{y}(t)$, so that $S = S(t, x(t))$ and conversely, $x(t) = X(t, S(t))$.
- If these functions $\bar{y}(t)$ and $X(t,s)$ were known, then the evaluation of the conditional expectation (58) would boil down to

$$\varphi(t,s) = \frac{\partial S}{\partial x}(t, X(t,s), \bar{y}(t)) \cdot \sigma_r(t, X(t,s), \bar{y}(t))$$

  where $\frac{\partial S}{\partial x}(t,x,y)$ is given by (57).

Once the volatility function $\varphi(t,s)$ is determined, namely once we have approximated $\bar{y}(t)$ and $X(t,s)$, swaption values can be computed from (58) by using methods for local volatility vanilla models (c.f. [AP10-1, Section 7.2]). We summarize these approximations in the following proposition (see [AP10-2, Section 13.1.4] for details).

Remark 6.1 (A simple approximation for low volatilities)**.** Setting $\bar{y}(t) \equiv 0$ and applying a linear approximation

$$S(t,x,0) = S(t,0,0) + \frac{\partial S}{\partial x}(t,0,0)x, \qquad \frac{\partial S}{\partial x}(t,x,0) = \frac{\partial S}{\partial x}(t,0,0)$$

whence

$$x \simeq \frac{S(t,x,0) - S(t,0,0)}{\frac{\partial S}{\partial x}(t,0,0)}$$

and we obtain

$$\varphi(t,s) \simeq \frac{\partial S}{\partial x}(t,0,0)\sigma_r\left(t, \frac{s - S(t,0,0)}{\frac{\partial S}{\partial x}(t,0,0)}, 0\right)$$

Proposition 6.3 (approximations of $y(t)$ and $X(t,s)$)**.** *(i) [AP10-2, Proposition 13.1.4] We approximate $y(t)$ by*

$$\begin{aligned} \bar{y}(t) &\simeq E^A[y(t)] = h^2(t)\int_0^t \left[\frac{\sigma_r(s,0,0)}{h(s)}\right]^2 ds, \quad t \in [0, T_0] \\ h(t) &= \exp\left(-\int_0^t \kappa(u)\,du\right) \end{aligned} \tag{60}$$

(ii) [AP10-2, Lemma 13.1.6] We approximate $x(t)$ by

$$\bar{x}(t) = E^A[x(t)] = x_0(t) + \frac{1}{2}\frac{\partial^2 X}{\partial s^2}(t, S(0))\, Var^A(S(t)), \quad t \in [0, T_0] \tag{61}$$

where $x_0(t)$ is given as the solution of the equation

$$S(t, x_0(t), \bar{y}(t)) = S(0)$$

which can be solved in a few iterations of Newton's algorithm (staring the search with $x = 0$), and where the variance $Var^A(S(t))$ can be approximated by

$$Var^A(S(t)) \simeq \int_0^t \left[\frac{\partial S}{\partial x}(s,0,0)\sigma_r(s,0,0)\right]^2 ds$$

The derivative $\frac{\partial^2 X}{\partial s^2}(t,x)$ in (61) can be computed by differentiating the implicit definition $S(t, X(t,s), \bar{y}(t)) \equiv s$ twice[13].

[13]Namely,

$$\frac{\partial X}{\partial s}(t,s) = \left(\frac{\partial S}{\partial x}(t,\right)^{-1} \qquad \text{and} \qquad \frac{\partial^2 X}{\partial s^2} = \frac{\frac{\partial^2 S}{\partial x^2}}{\left(\frac{\partial S}{\partial x}\right)^2}$$

(iii) *[AP10-2, Proposition 13.1.8] In order to approximate $X = X(t,s)$, since $S(t,x)$ is observed to be closely approximated as a quadratic function of x, we perform a second order expansion of*

$$S(t, X(t,s), \bar{y}(t)) = s$$

around $\bar{x}(t)$, namely we approximate $X(t,s)$ with the solution $\xi = \xi(t,s)$ of the quadratic equation

$$S(t,\bar{x}(t),\bar{y}(t)) + \frac{\partial S}{\partial x}(t,\bar{x}(t),\bar{y}(t))[\xi - \bar{x}(t)] + \frac{1}{2}\frac{\partial^2 S}{\partial x^2}(t,\bar{x}(t),\bar{y}(t))[\xi - \bar{x}(t)]^2 = s \quad (62)$$

(iv) *In conclusion, the conditional expectation (59)*

$$\varphi^2(t,s) = E^A\left[\left\{\frac{\partial S}{\partial x}(t,x(t),y(t))\sigma_r(t,x(t),y(t))\right\}^2 |S(t) = s\right]$$

can be approximated by

$$\varphi(t,s) \approx \frac{\partial S}{\partial x}(t,\xi(t,s),\bar{y}(t)) \cdot \sigma_r(t,\xi(t,s),\bar{y}(t)) \quad (63)$$

where

- *$\frac{\partial S}{\partial x}(t,x,y)$ is given by (57).*
- *$\frac{\partial^2 X}{\partial s^2}(x,t)$ can be computed by differentiating the implicit definition $S(t,X(t,s),\bar{y}(t)) \equiv s$ twice.*
- *$\xi(t,s)$ is the solution of the quadratic equation[14] (62),*
- *$\bar{x}(t)$ is given by (61),*
- *$\bar{y}(t)$ is given by (60),*

6.2.1 Linear local volatility

Even though one could use the previous results to calibrate the function $\sigma_r(t,x,y)$ non-parametrically to the implied volatilities of a collection of swaptions across all strikes, it is recommendable to chose $\sigma_r(t,x,y)$ from a parametric family of monotone, downward sloping functions of state variables.

- We next make the **assumption that the short rate local volatility function is linear**

  $$\sigma_r(t,x,y) = \lambda_r(t)\left[\alpha_r(t) + b_r(t)x\right] \quad (64)$$

  where the scale function $\alpha_r(t)$ is fixed exogenously and the volatility $\lambda_r(t)$ and skew $b_r(t)$ are calibrated to the market. The swap rate local volatility function $\varphi(t,s)$ (63) thus becomes

  $$\varphi(t,s) \approx \frac{\partial S}{\partial x}(t,\xi(t,s),\bar{y}(t)) \cdot \lambda_r(t)\left[\alpha_r(t) + b_r(t)\xi(t,s)\right]$$

- We further **assume that the swap rate local volatility function $\varphi(t,s)$ is well approximated by a linear function** in s too, which leads to the following proposition [AP10-2, Corollary 13.1.9]:

[14] Any of the 2 solutions?

Proposition 6.4 (approximate swap rate dynamics with linear local volatility function)**.** *The swap rate local volatility function $\varphi(t, S(t))$ can be approximated by*

$$
\begin{aligned}
\varphi(t,S(t)) &\approx \varphi(t,S(0)) + \frac{\partial \varphi}{\partial s}(t,S(0))[S(t)-S(0)] \\
&= \varphi(t,S(0)) \left[\frac{\frac{\partial \varphi}{\partial s}(t,S(0))}{\varphi(t,S(0))} S(t) + \left(1 - \frac{\frac{\partial \varphi}{\partial s}(t,S(0))}{\varphi(t,S(0))}\right) S(0) \right] \\
&\overset{not}{=} \lambda_S(t) \left[b_S(t) S(t) + (1 - b_S(t)) S(0) \right]
\end{aligned}
$$

where

$$
\lambda_S(t) = \lambda_r(t) \frac{1}{S(0)} \frac{\partial S}{\partial x}(t, \bar{x}(t), \bar{y}(t)) \left[\alpha_r(t) + b_r(t)\bar{x}(t) \right] \tag{65}
$$

$$
b_S(t) = \frac{S(0)}{\alpha_r(t) + b_r(t)\bar{x}(t)} \frac{b_r(t)}{\frac{\partial S}{\partial x}(t, \bar{x}(t), \bar{y}(t))} + \frac{S(0) \frac{\partial^2 S}{\partial x^2}(t, \bar{x}(t), \bar{y}(t))}{\left[\frac{\partial S}{\partial x}(t, \bar{x}(t), \bar{y}(t))\right]^2} \tag{66}
$$

The dynamics of the swap rate (58) are hence approximated by

$$
dS(t) = \varphi(t, S(t)) dW^A(t) = \lambda_S(t) \left[b_S(t) S(t) + (1 - b_S(t)) S(0) \right] dW^A(t) \tag{67}
$$

The model (67) is a displaced log-normal SDE with time-dependent volatility $\lambda_S(t)$ and skew $b_S(t)$ and it can be converted into a displaced log-normal SDE with time-constant coefficients using averaging methods. This yields the following approximate swaption pricing formula [AP10-2, Proposition 13.1.10].

Proposition 6.5 (approximate swaption pricing in local volatility qG model)**.** *Consider a payer swaption with strike c and expiry T_0 on the swap rate $S(t)$. In the local-volatility quasi-Gaussian model with linear short rate volatility*

$$
\sigma_r(t, x, y) = \lambda_r(t) \left[\alpha_r(t) + b_r(t) x \right],
$$

the swaption price can be approximated by the displaced log-normal option formula

$$
\begin{aligned}
V_{swaption}(0) &\simeq A(0) \left[\left(S(0) + S(0) \frac{1 - \bar{b}_S}{\bar{b}_S} \right) \mathcal{N}(d_+) - \left(c + S(0) \frac{1 - \bar{b}_S}{\bar{b}_S} \right) \mathcal{N}(d_-) \right] \\
d_\pm &= \frac{\ln\left(\frac{S(0) + S(0) \frac{1 - \bar{b}_S}{\bar{b}_S}}{c + S(0) S(0) \frac{1 - \bar{b}_S}{\bar{b}_S}} \right) \pm \frac{1}{2} \bar{b}_S^2 \bar{\lambda}_S^2 T_0}{\bar{b}_S \bar{\lambda}_S \sqrt{T_0}}
\end{aligned}
$$

where

$$
\bar{\lambda}_S = \left(\frac{1}{T_0} \int_0^{T_0} \lambda_S^2(t)\, dt \right)^{1/2} \tag{68}
$$

$$
\bar{b}_S = \int_0^{T_0} b_S(t) \omega_S(t)\, dt \tag{69}
$$

$$
\omega_S(t) = \frac{\lambda_S(t)^2 \int_0^t \lambda_S(s)^2\, ds}{\int_0^{T_0} \left(\lambda_S(u)^2 \int_0^u \lambda_S(s)^2\, ds \right) du} \tag{70}
$$

6.2.2 Volatility calibration

For calibration purposes, we assume that the parameters $\alpha_r(t), b_r(t), \lambda_r(t)$ defining the local volatility function (64) of the qG model are piecewise constant on a maturity grid.

Concretely, given $0 = T_0 < T_1 < \cdots < T_{N-1}$:

- Consider a collection of $N-1$ coterminal options, with the n-th one expiring at T_n.
- Assume the n-th swaption has an underlying swap with $\mu(n)$ periods and denote the corresponding swap rate and annuity

$$S_n(t) \equiv S_{n,\mu(n)}(t), \qquad A_n(t) \equiv A_{n,\mu(n)}(t), \quad n = 1, \ldots, N-1$$

- The parameter $\alpha_r(t)$ is redundant, so we elect to set $\alpha_r(t) = S_n(0) 1_{(T_{n-1}, Tn]}(t)$.

With these observations in mind, the local volatility definition (64) specializes to

$$\sigma_r(t, x, y) = \sum_{n=1}^{N-1} \lambda_n \left[S_n(0) + b_n D_n x \right] 1_{(T_{n-1}, Tn]}(t), \qquad D_n = \frac{\partial S_n}{\partial x}(t, 0, 0)$$

where

$$\lambda_r(t) = \sum_{n=1}^{N-1} \lambda_n 1_{(T_{n-1}, Tn]}(t), \quad \alpha_r(t) = \sum_{n=1}^{N-1} \alpha_n 1_{(T_{n-1}, Tn]}(t), \quad b_r(t) = \sum_{n=1}^{N-1} b_n 1_{(T_{n-1}, Tn]}(t)$$

Assume for now that the mean reversion function $\kappa(t)$ is specified externally and that a collection of market parameters $(\hat{\lambda}_{S_n}, \hat{b}_{S_n})$, $n = 1, \ldots, N-1$ is given[15]. From Proposition 6.5, the value of a swaption expiring at T_n depends only on parameters (λ_i, b_i) for $i = 1, \ldots, n$, so the local volatility qG model can be calibrated by a bootstrap method as follows:

1. Starting values: set the λ_n's to volatilities obtained by calibrating a pure Gaussian model and $b_n = \hat{b}_{S_n}$, $n = 1, \ldots, N-1$. Set $n = 1$.
2. Given n, parameters (λ_i^*, b_i^*) are already calibrated for $i = 1, \ldots, n-1$. Use (61) and (60) to update $\bar{x}(t)$ and $\bar{y}(t)$ for $t \in [0, T_n]$ using the former, together with the starting value (λ_n, b_n).
3. Calculate $\lambda_{S_n}(t)$ and $b_{S_n}(t)$ for $t \in [0, T_{n-1}]$ using (65) and (66) respectively.
4. Make another guess for (λ_n, b_n) and use it to update $\lambda_{S_n}(t)$ and $b_{S_n}(t)$ for $t \in [T_{n-1}, T_n]$, once again using (65) and (66) respectively.
5. Calculate the averaged parameters $\bar{\lambda}_{S_n}$ and $\bar{b}_{S_n}$ using (68) and (69) from Proposition 6.5.
6. If $(\bar{\lambda}_{S_n}, \bar{b}_{S_n})$ is not equal to $(\hat{\lambda}_{S_n}, \hat{b}_{S_n})$ within a given tolerance, go to Step 4.
   Essentially, in this step we seek to obtain (λ_n, b_n) such that

$$\left\| (\bar{\lambda}_{S_n} - \hat{\lambda}_{S_n}, \bar{b}_{S_n} - \hat{b}_{S_n}) \right\|^2 \leq \epsilon$$

   for a given tolerance ϵ. Once accomplished, move to the next step.
7. Set $(\lambda_n^*, b_n^*) = (\lambda_n, b_n)$, update $n \mapsto n+1 \leq N-1$ and go to Step 2.

[15] These can be obtained by fitting a series of constant parameter displaced log-normal vanilla models to the observed swaption volatility smiles at all expiries.

6.2.3 Computational tip

We explain in detail how to perform the n-th step...

$\bar{y}(t)$

$\bar{x}(t)$

6.2.4 Mean reversion calibration

The calibration of the mean reversion speed $\kappa(t)$ of the local volatility qG model relies on the observation that the ratio of volatilities of two swaptions with the same expiry date is practically independent of volatility. This suggests using a second strip of swap rates to define targets for mean reversion calibration as the ratios between these rates and the original ones.

Consider the original strip of swap rates $\{S_{n,\nu(n)}(\cdot)\}$, $n = 1, \ldots, N-1$ and obtain a new one $\{S_{n,\mu(n)}(\cdot)\}$, $n = 1, \ldots, N-1$ with identical fixing dates, but spanning possibly different periods.

Proposition 6.6. *Consider the quasi-Gaussian model with local short rate volatility function $\sigma_r(t) = g(t, x(t), y(t))h(t)$. The ratio of two swap rates fixing on T_n with m_1 and m_2 periods, respectively, is approximately given by*

$$\frac{Var(S_{n,m_1}(T_n))}{Var(S_{n,m_2}(T_n))} \approx \frac{\int_0^{T_n} \left(\frac{\partial S_{n,m_1}}{\partial x}(t,0,0)\right)^2 dt}{\int_0^{T_n} \left(\frac{\partial S_{n,m_2}}{\partial x}(t,0,0)\right)^2 dt} \tag{71}$$

Note that the ratios (71) depend on the mean reversion parameter $\kappa(t)$ only, so that it can be calibrated independently of the other parameters. For calibration we add the usual piecewise constant assumption

$$\kappa(t) = \sum_{n=1}^{N-1} \kappa_n 1_{(T_{n-1},T_n]}(t) + \kappa_N 1_{(T_{N-1},\infty]}(t)$$

We then set the optimal mean reversion levels $\{\kappa_n\}$ as the solution of the optimization problem

$$\{\kappa_n^*\} = \arg\min \left\{ \sum_{n=1}^{N-1} \left[\frac{\mathrm{Var}(S_{n,m_1}(T_n))}{\mathrm{Var}(S_{n,m_2}(T_n))}(\{\kappa_n\}) - \frac{\widehat{\mathrm{Var}}(S_{n,m_1}(T_n))}{\widehat{\mathrm{Var}}(S_{n,m_2}(T_n))} \right]^2 + \omega \sum_{n=1}^{N-1} (\kappa_{n+1} - \kappa_n)^2 \right\}$$

where we included a regularization term penalizing non-stationary behavior (dependent on a user-specified weight ω) and where the $\widehat{Var}(S_{n,m})$'s are market-implied variances of swap rates, which in the case of the linear short rate volatility model under consideration can be approximated[16] by

$$\widehat{Var}(S_{n,m}(T_n)) \approx \left(S_{n,m}(0)\hat{\lambda}_{S_{n,m}}\right)^2 T_n$$

6.3 Stochastic volatility

Local volatility models are characterized by a dependence of the function g in the factorization $\sigma_f(t, T, w) = g(t, w)h(T)$ on the state variables, namely $g(t) \equiv g(t, x(t), y(t))$. Instead,

[16] Market-implied variances can be calculated in more general models from values of options on the swap rates.

we can enrich the model by incorporating an additional stochastic factor via following specification

$$
\begin{aligned}
g(t,w) &\equiv \sqrt{z(t)}g(t,x(t),y(t)) && (72)\\
dz(t) &= \theta[z_0 - z(t)]dt + \eta(t)\sqrt{z(t)}dZ(t), && (73)\\
E[dW(t),dZ(t)] &= 0
\end{aligned}
$$

The analysis from the previous section carries through similarly in this context. In particular, the model is defined by the collection of SDE's

$$
\begin{aligned}
dx(t) &= [y(t) - \kappa(t)x(t)]\,dt + \sigma_r(t,x(t),y(t))dW(t), \quad x(0)=0,\\
dy(t) &= \left[\sigma_r^2(t,x(t),y(t)) - 2\kappa(t)y(t)\right]dt, \quad y(0)=0,\\
dz(t) &= \theta(z_0 - z(t))dt + \eta(t)\sqrt{z(t)}dZ(t), \quad z(0)=z_0=1,\\
dZ(t)dW^A(t) &= 0
\end{aligned}
$$

A few remarks are in order:

- The bond reconstitution formula (55) is the same as for any other quasi-Gaussian model (c.f. Lemma 6.1.
- In particular, the bond reconstitution formulas do not depend on the stochastic volatility process $z(t)$, so the SV-qG model is a true stochastic volatility model, in the sense that the stochastic volatility is unspanned and cannot be hedged by discount bonds.
- The time-dependent volatility of variance is assumed to be piecewise constant $\eta(t) = \sum_{n=1}^{N-1} \eta_n \mathbb{I}_{(T_{n-1},T_n]}(t)$.

An analysis identical to the one carried out in the local volatility setting yields the following approximate dynamics for the short rate:

Proposition 6.7 (approximate dynamics for the swap rate $S(t)$ in the SV-qV model)**.** *Assuming that the local short rate volatility is linear $\sigma_r(t,x,y) = \lambda_r(t)\left[\alpha_r(t) + b_r(t)x\right]$, the dynamics of the swap rate $S(t)$ in the stochastic volatility quasi-Gaussian model (**??**) are given approximately by*

$$
\begin{aligned}
dS(t) &= \sqrt{z}(t)\lambda_S(t)\left[b_S(t)S(t) + (1-b_S(t))S(0)\right]dW^A(t),\\
dz(t) &= \theta(z_0 - z(t))dt + \eta(t)\sqrt{z(t)}dZ(t), \quad z(0)=z_0=1,\\
dZ(t)dW^A(t) &= 0
\end{aligned}
$$

where $\lambda_S(t)$ and $b_S(t)$ are given by equations (65)-(66).

This is a stochastic volatility model with time-dependent parameters for which the time averaging methods of [AP10-1, Sction 9.3] are available.

6.4 Quasi-Gaussian multi-factor models

7 The Libor market model

The **Libor market model** (LM) is a model capable of:

- Capturing the full correlation structure of across the entire yield curve.
- Allowing volatility calibration to a large enough set f European options that the volatility characteristics of most exotic securities can be considered *spanned* by the calibration.

7.1 The basics

The **HJM framework** generally involves multiple driving Brownian motions and an infinite set of state variables (the instantaneous forward rates). Working directly with forward rates is unattractive for a number of reasons:

(i) Instantaneous forward rates are never quoted in the market: realistic securities involve simply compounded (Libor) rates. The development of market-consistent pricing expressions is thus cumbersome.

(ii) An infinite set of instantaneous forward rates will require discretization.

(iii) Prescribing the form of the volatility function of instantaneous forward rates is complicated.

BGM, Jamshidian: All three issues can be solved by **formulating the model in terms of a non-overlapping set of simply compounded Libor rates**.

Fix a tenor structure

$$0 = T_0 < T_1 < \cdots < T_N, \qquad \tau_n = T_n - T_{n-1}$$

Instead of keeping track of an entire yield curve, at any point in time we are focused only on a finite set of zero-coupon bonds

$$P(t, T_n), \qquad t < T_n \leq T_N$$

Define $q(t)$ to be the tenor structure index of the shortest-dated discount bond still alive at time t, namely

$$T_{q(t)-1} \leq t < T_{q(t)}$$

We define the **Libor forward rates** as

$$L_n(t) = L(t, T_n, T_{n+1}) = \frac{1}{\tau_n}\left(\frac{P(t,T_n)}{P(t,T_{n+1})} - 1\right), \qquad q(t) \leq n \leq N-1$$

7.1.1 Probability measures

By construction, $L_n(t)$ is a $Q^{T_{n+1}}$-martingale, so by the martingale representation theorem there is an adapted vector process $\sigma_n(t)$ such that

$$\boxed{dL_n(t) = \sigma_n(t)^\top dW^{n+1}(t)}$$

In order towrite the process in measure Q^{T_n}, consider the Radon-Nikodym process

$$\xi(t) = \left(\frac{dQ^{T_n}}{dQ^{T^{n+1}}}\right)_t = (1 + \tau_n L_n(t))\frac{P(0,T_{n+1})}{P(0,T_n)}, \quad \frac{d\xi(t)}{\xi(t)} = \frac{\tau_n \sigma_n(t)^\top dW^{n+1}(t)}{1 + \tau_n L_n(t)}$$

From Girsanov's theorem, $dW^n(t) = dW^{n+1}(t) - \frac{\tau_n \sigma_n(t)^\top}{1+\tau_n L_n(t)} dt$ is a Q^{T_n}-BM and

$$\boxed{dL_n(t) = \sigma_n(t)^\top \left(\frac{\tau_n \sigma_n(t)}{1+\tau_n L_n(t)} dt + dW^n(t) \right)}$$

Iterating this argument, one easily sees that in the terminal measure Q^{T_N}, the process for $L_n(t)$ is

$$\boxed{dL_n(t) = \sigma_n(t)^\top \left(-\sum_{j=n+1}^{N-1} \frac{\tau_j \sigma_j(t)}{1+\tau_j L_j(t)} dt + dW^N(t) \right)}$$

As for the spot measure, recall that it is the measure induced by taking as numéraire

$$B(t) = P(t, T_{q(t)}) \prod_{n=0}^{q(t)-1} (1 + \tau_n L_n(T_n))$$

The random part of the numéraire is the discount bond $P(t, T_{q(t)})$, so the dynamics in measure Q^B coincide with the dynamics of $Q^{T_{q(T)}}$ and arguing as above we see

$$\boxed{dL_n(t) = \sigma_n(t)^\top \left(\sum_{j=q(t)}^{n} \frac{\tau_j \sigma_j(t)}{1+\tau_j L_j(t)} dt + dW^B(t) \right)}$$

7.1.2 Link to HJM

Recall that HJM models have risk-neutral dynamics of the form

$$df(t,T) = \sigma_f(t,T)^\top \int_t^T \sigma_f(t,u)\, du dt + \sigma_f(t,T)^\top dW(T)$$

The dynamics for the forward bond $P(t, T_n, T_{n+1})$ are of the form

$$\frac{dP(t,T_n,T_{n+1})}{P(t,T_n,T_{n+1})} = \mathcal{O}(dt) + \left(\sigma_P(t,T_{n+1})^\top - \sigma_P(t,T_n)^\top \right) dW(t)$$

By the definition of the forward Libor rate $L_n(t) = \frac{1}{\tau_n} \left(\frac{P(t,T_n)}{P(t,T_{n+1})} - 1 \right)$ and using the diffusion invariance principle one sees that

$$\sigma_n(t) = \frac{1}{\tau_n} (1 + \tau_n L_n(t)) \int_{T_n}^{T_{n+1}} \sigma_f(t,u)\, du$$

We thus see that the LM model is a special case of the HJM model: a full specification of $\sigma_f(t,T)$ uniquelt determined the LM volatility $\sigma_n(t)$, but the converse is not true.

7.1.3 Choice of $\sigma_n(t)$

To build a workable model, we need to be specific about $\boldsymbol{\sigma}_n(t)$.

(i) Local volatility LM

$$\boldsymbol{\sigma}_n(t) = \boldsymbol{\lambda}_n(t) \varphi(L_n(t))$$

where $\boldsymbol{\lambda}_n(t)$ is a bounded vector-valued deterministic function and φ is a time-homogeneous local volatility function.

(ii) Stochastic volatility LM: we extend the previous local volatility model and we allow the term on the Brownian motion to be scaled by a stochastic process so we have

$$\begin{aligned} dL_n(t) &= \sqrt{z(t)}\varphi(L_n(t))\boldsymbol{\lambda}_n(t)^\top \left(\sqrt{z(t)}\boldsymbol{\mu}_n(t)dt + d\boldsymbol{W}^{\boldsymbol{B}}(t)\right) && (74) \\ \boldsymbol{\mu}_n(t) &= \sum_{j=q(t)}^{n} \frac{\tau_j\varphi(L_j(t))\boldsymbol{\lambda}_j(t)}{1+\tau_j L_j(t)} \\ dz(t) &= \theta[z_0 - z(t)]dt + \eta\psi(z(t))dZ(t), \quad z(0) = 0 && (75) \end{aligned}$$

- Note that a single common factor $\sqrt{z(t)}$ simultaneously scales all forward rate volatilities (even though this can be relaxed, see section ??).
- In order to make the variance process (75) more tractable, it is common to assume that the Brownian motions $Z(t)$ and $W^B(t)$ are uncorrelated.

7.2 Correlation structure

In one-factor models, where all the points on the forward curve move in the same direction. Empirical evidence contradicts this and indicates that various points on the forward curve do not move co-monotonically with each other. The LM model gives us control over the instantaneous correlation between various points, via the simple observation

$$Corr(dL_k(t), dL_j(t)) = \frac{\boldsymbol{\lambda}_k(t)^\top \boldsymbol{\lambda}_j(t)}{\|\boldsymbol{\lambda}_k(t)\|\|\boldsymbol{\lambda}_j(t)\|}$$

As we add more Brownian motions, our ability to capture increasingly complicated correlation structures progressively improves, at the cost of increasing the model complexity. In order to choose the model dimension m we may follow two approaches:

(i) Fix the correlation structure in the model to match empirical forward rate correlations, for instance via PCA.

(ii) Imply the correlation structure directly from traded market data by amending the set of calibration instruments with securities that have stronger sensitivity to forward rate correlations, like *yield curve spread options.*

7.2.1 Empirical PCA

For fixed τ (e.g. 0.25 or 0.5), define the rates

$$\ell(t,x) = L(t, t+x, t+\tau)$$

Fix a set of tenors $x_1, \ldots, x_{N_x}$ and a set of calendar times $t_1, \ldots, t_{N_t}$ and set up the $N_x \times N_t$ observation matrix

$$O_{i,j} = \frac{\ell(t_j, x_i) - \ell(t_{j-1}, x_i)}{\sqrt{t_j - t_{j-1}}}, \quad i = 1, \ldots, N_x, \quad j = 1, \ldots, N_t$$

where for each date in the time grid t_j, we construct the forward curve (starting at t_j) from market observable swaps, futures and deposits. Use this to construct a variance covariance matrix

$$C = \frac{OO^\top}{N_t}$$

In order for the LM model to conform to empirical data, *we need sufficiently many driving Brownian motions m to replicate this variance matrix.* This is accomplished by performing a PCA: compute the eigenvalues of C together with the percentage of variance that they explain, and decide how many are necessary to explain a reasonable amount thereof (generally 4 or 5).

7.2.2 Empirical estimates for forward rate correlations

Introduce the diagonal matrix

$$c = \operatorname{diag}\left(\sqrt{C_{1,1}}, \ldots, \sqrt{C_{N_x,N_x}}\right)$$

so that the *empirical forward rate correlation matrix* becomes

$$R = c^{-1} C c^{-1}. \tag{76}$$

The term $R_{i,j}$ provides a sample estimate of the instantaneous correlation between increments in $\ell(t, x_i)$ and $\ell(t, x_j)$

$$R_{i,j} = \frac{1}{\sqrt{C_{i,i}}\sqrt{C_{j,j}}} C_{i,j} = \frac{1}{\sqrt{C_{i,i}}\sqrt{C_{j,j}}} \frac{1}{N_t} \sum_{k=1}^{N_t} \frac{[\ell(t_k, x_i) - \ell(t_{k-1}, x_i)] \cdot [\ell(t_k, x_j) - \ell(t_{k-1}, x_j)]}{t_k - t_{k-1}}.$$

In multi-factor yield curve modeling, it is common practice to work with simple parametric forms for correlations

$$\operatorname{Corr}(L_k(t), L_j(t)) = q(T_k - t, T_j - t)$$

For instance

$$\begin{aligned} q(x,y) &= \rho_\infty(\min\{x,y\}) + [1 - \rho_\infty(\min\{x,y\})]\, e^{-a(\min\{x,y\}) \cdot |y-x|}, \qquad (77) \\ a(z) &= a_\infty + (a_0 - a_\infty)^{\kappa z}, \\ \rho_\infty(z) &= b_\infty + (b_0 - b_\infty) e^{-\alpha z} \end{aligned}$$

where $\xi = (a_0, a_\infty, \kappa, b_0, b_\infty, \alpha)^\top$ is a vector of parameters that are found by least-squares optimization against an empirical correlation matrix (76). Concretely, if $R_2(\xi)$ is the correlation matrix generated from the form (77), the optimal parameter vector ξ^* is obtained by minimizing

$$\xi^* = \arg\min_\xi \left[\operatorname{tr}\left((R - R_2(\xi)) \cdot (R - R_2(\xi))^\top\right)\right], \quad \xi \geq 0$$

7.3 Pricing European options

In order to calibrate the model

$$\begin{aligned} dL_n(t) &= \sqrt{z(t)} \varphi(L_n(t)) \boldsymbol{\lambda}_n(t)^\top \left(\sqrt{z(t)} \boldsymbol{\mu}_n(t) dt + d\boldsymbol{W}^B(t)\right) \\ \boldsymbol{\mu}_n(t) &= \sum_{j=q(t)}^{n} \frac{\tau_j \varphi(L_j(t)) \boldsymbol{\lambda}_j(t)}{1 + \tau_j L_j(t)} \\ dz(t) &= \theta[z_0 - z(t)]dt + \eta \psi(z(t)) dZ(t), \quad z(0) = 0 \end{aligned}$$

the vectors $\boldsymbol{\lambda}_k(t)$ must be such that the model successfully reproduces the prices of liquid plain-vanilla derivatives (i.e. swaptions and caps). We thus need to find pricing formulas for vanilla options that are fast enough to be embedded into an iterative calibration algorithm.

7.3.1 Caplets

An *interest rate cap* is a security that allows one to benefit from low floating rates, yet be protected from high rates. This is accomplished by giving the holder the right to pay the smaller of two simple rates: the floating rate f and a fixed rate κ: the holder of a cap over the interval $[T, T+\tau]$, the interest rate paid at time $T + \tau$ on each dollar of principal is

$\min\{\tau\kappa, \tau f\}$. Without the caplet the interest payment would be τf, so the caplet's worth to the holder is

$$\tau f - \min\{\tau f, \tau\kappa\} = (f-\kappa)^+\tau$$

Given a tenor structure $T_0 < T_1 < \cdots < T_N$ and forward Libor rates $L_n(t) = L(t; T_n, T_{n+1})$, a cap is a stream of κ-caplets maturing at times T_n and paying out

$$V_{caplet}(T_{n+1}) = \tau_n \left(L_n(T_n) - c\right)^+$$

The time-t value of a caplet is hence

$$V_{caplet}(t) = B(t)E_t\left[\frac{1}{B(T_{n+1})}V_{caplet}(T_{n+1})\right] = B(t)E_t\left[\frac{1}{B(T_{n+1})}\tau_n \left(L_n(T_n) - c\right)^+\right]$$

and the time-t value of a cap in the both the risk-neutral is

$$V_{cap}(t) = B(t)\sum_{n=0}^{N-1} E_t\left[\frac{1}{B(T_{n+1})}\tau_n \left(L_n(T_n) - c\right)^+\right]$$

Switching to the T_{n+1}-forward measure[17] for the n-th caplet, the pricing formula becomes

$$V_{cap}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) E_t^{T_{n+1}}\left[(L_n(T_n) - c)^+\right]$$

Remark 7.1. As can be seen from the last expression, *the terminal correlation between the forward rates L_n does not affect the cap value.* Unfortunately, this is no longer the case for swaptions.

Proposition 7.1 (Caplet pricing in the SV-LM model (74)-(75))**.** *Assume that the forward rate dynamics in the spot measure are given by (74)-(75). Then*

$$V_{caplet}(0) = P(0, T_{n+1})\tau_n E^{T_{n+1}}\left[(L_n(T_n) - c)^+\right] \tag{78}$$

where

$$\begin{aligned} dL_n(t) &= \sqrt{z(t)}\varphi(L_n(t))\|\lambda_n(t)\|dY^{n+1}(t) \\ dz(t) &= \theta[z_0 - z(t)]dt + \eta\psi(z(t))dZ(t) \\ Y^{n+1}(t) &= \int_0^t \frac{\lambda_n(s)^\top}{\|\lambda_n(s)\|}\, dW^{n+1}(s) \end{aligned}$$

We are thus reduced to computing the expectation of $(L_n(T_n) - c)^+$, where the process $L_n(t)$ is a standard scalar stochastic volatility diffusion.

7.3.2 Swaptions

Recall that a fixed-for-floating interest rate swap allows to exchange a stream of cash-flows at a floating (Libor) rate by a stream of fixed-rate cash-flows. Given a tenor structure $0 \le T_0 < T_1 < \cdots < T_N$, from the perspective of the fixed-rate payer, a swap pays out

$$V_{swap}(T_{n+1}) = \tau_n(L_n(T_n) - \kappa)$$

[17] Recall that the pricing formula with numéraire g reads

$$V(t) = g(t)E\left[\frac{1}{g(T)}V(T)|\mathcal{F}_t\right]$$

The time-t price of a swap is hence given by

$$\begin{aligned}
V_{swap}(t) &= \beta(t)\sum_{n=0}^{N-1} E_t\left[\frac{1}{\beta(T_{n+1})}\tau_n(L_n(T_n)-\kappa)\right] \\
&\stackrel{[1]}{=} \sum_{n=0}^{N-1}\tau_n P(t,T_{n+1})(L_n(t)-\kappa) \\
&\stackrel{[2]}{=} A(t)\,(S(t)-\kappa)
\end{aligned}$$

where:

- Equality [1] shows that a vanilla fixed-floating swap can be valued on date t using only the term structure of interest rates observed on that date. In order to obtain [1], one starts by using that $P(T_n,T_{n+1}) = E\left[\frac{\beta(T_n)}{\beta(T_{n+1})}\right]$, then the definition $1+\tau_n L_n(t) = \frac{P(t,T_n)}{P(t,T_{n+1})}$ and finally the fact that $\frac{P(t,T_n)}{\beta(t)}$ is a martingale.
- In [2] we denoted

$$A(t) = A_{0,N}(t) = \sum_{n=0}^{N-1}\tau_n P(t,T_n), \qquad S(t) = S_{0,N}(t) = \frac{1}{A(t)}\sum_{n=0}^{N-1}\tau_n P(t,T_n)L_n(t)$$

A *European swaption* is a contract giving the holder the right, but no the obligation, to enter a swap at a future date, at a given fixed rate. A *payer* (*receiver*) swaption is an option to pay (receive) the fixed leg on a fixed-floating swap.

The payoff for a payer swaption maturing at time T_j, with the underlying security being a fixed-for-floating swap making payments at times $T_{j+1},\ldots,T_k$, $j<k\le N$ is simply

$$V_{swaption}(T_j) = (V_{swap}(T_j))^+ = \left(\sum_{n=j}^{k}\tau_n P(T_j,T_{n+1})(L_n(T_0)-\kappa)\right)^+$$

and for $t<T_j$, the time-t value of a payer swaption is given by

$$\begin{aligned}
V_{swaption}(t) &= \beta(t)E_t\left[\frac{1}{\beta(T_j)}V_{swaption}(T_j)\right] \\
&= \beta(t)E_t\left[\left(\frac{1}{\beta(T_j)}\sum_{n=j}^{k-1}\tau_n P(T_j,T_{n+1})(L_n(T_0)-\kappa)\right)^+\right] \\
&= \beta(t)E_t\left[\frac{1}{\beta(T_j)}A(T_j)(S(T_j)-\kappa)^+\right] \\
&= A(t)E_t^A\left[(S(T_j)-\kappa)^+\right]
\end{aligned}$$

where

$$A(t) = A_{j,k-j}(t) = \sum_{n=j}^{k-1}P(t,T_{n+1})\tau_n, \qquad S(t) = S_{j,k-j}(t) = \frac{P(t,T_j)-P(t,T_k)}{A(t)}$$

Pricing swaptions in the Libor model is more involved than pricing caps and will require some amount of approximation if a quick algorithm is required. Since $S(t)$ is clearly a Q^A-martingale, an application of Itô's lemma easily yields the following:

Lemma 7.2 (Swaption pricing in the SV-LM model (74)-(75)). *Assume that the forward rate dynamics in the spot measure are given by (74)-(75) and let Q^A be the measure induced by using $A(t)$ as numéraire. Then*

$$dS(t) = \sqrt{z(t)}\varphi(S(t)) \sum_{n=j}^{k-1} w_n(t)\lambda_n(t)^\top dW^A(t) \tag{79}$$

where

$$\begin{aligned} w_n(t) &= \frac{\varphi(L_n(t))}{\varphi(S(t))} \frac{\partial S(t)}{\partial L_n(T)} \\ &= \frac{\varphi(L_n(t))}{\varphi(S(t))} \times \frac{S(t)\tau_n}{1+\tau_n L_n(t)} \times \left[\frac{P(t,T_k)}{P(t,T_j) - P(t,T_k)} + \frac{1}{A(t)} \sum_{i=n}^{k-1} \tau_i P(t, T_{i+1}) \right] \end{aligned} \tag{80}$$

Sketch of proof. Since $S(t)$ is Q^A-martingale, $dS(t)$ is driftless. Hence, by Itô's lemma, since $S(t) = S(L_j(t), \ldots, L_{k-1}(t))$, we have

$$dS(t) = \sum_{n=j}^{k-1} \frac{\partial S(t)}{\partial L_n(t)} \sqrt{z(t)}\varphi(L_n(t))\boldsymbol{\lambda}_n(t)^\top dW^A(t)$$

Evaluating the partial derivatives $\frac{\partial S(t)}{\partial L_n(t)}$, we obtain (79). □

The swap rate dynamics in (79) are clearly too complicated to allow analytical treatment. For simplicity, one generally assumes that the weights $w_n(t)$ are constant, and equal to their time-0 value[18].

Proposition 7.3 (Approximate swaption pricing in the SV-LM model (74)-(75)). *The time-0 price of the T_j-maturity swaption is given by*

$$V_{swaption}(0) = A(0)E^A\left[(S(T_j) - c)^+\right] \tag{81}$$

If $w_n(t)$ are the weights in (80), define

$$\lambda_S(t) = \sum_{n=j}^{k-1} w_n(0)\lambda_n(t)$$

The swap dynamics in Proposition 7.2 can be approximated by

$$\begin{aligned} dS(t) &\simeq \sqrt{z(t)}\varphi(S(t))\|\lambda_S(t)\|dY^A(t), && (82) \\ dz(t) &= \theta[z_0 - z(t)]dt + \eta\psi(z(t))dZ(t), \\ \|\lambda_S(t)\|dY^A(t) &= \sum_{n=j}^{k-1} w_n(0)\lambda_n(t)^\top dW^A(t) && (83) \end{aligned}$$

where $Y^A(t)$ and $Z(t)$ are independent scalar Brownian motions in measure Q^A.

- the scalar term $\|\lambda_S(t)\|$ is purely deterministic, so computation of the Q^A-expectation (81) can be obtained using standard analytical tools for scalar stochastic volatility models (c.f. Chapter 8 in [AP10-1]).

We next give an example for displaced log-normal Libor rates.

[18] C.f. discussion at the bottom of page 617 in [AP10-2].

Corollary 7.4 (swaption pricing under displaced log-normal Libor rates). *Assume that each rate $L_n(t)$ follows a displaced log-normal process in its own measure*

$$dL_n(t) = [bL_n(t) + (1-b)L_n(0)]\, \boldsymbol{\lambda}_n(t)^\top d\boldsymbol{W}^{n+1}(t), \quad n = 1, \ldots, N-1$$

the time-0 price of a T_j-maturity swaption is given by

$$V_{swaption,j}(0) = A(0)c_B\left(0, \frac{S(0)}{b}; T_j, c - S(0) + \frac{S(0)}{b}; b\bar{\lambda}_{S_j}\right)$$

with term swap rate volatility given by

$$\begin{aligned} \bar{\lambda}_{S_j} &= \frac{1}{T_j}\left(\int_0^{T_j} \|\lambda_S(t)\|^2\, dt\right)^{1/2}, \\ \lambda_S(t) &= \sum_{n=j}^{k-1} \omega_n(0)\lambda_n(t) \end{aligned}$$

Above $c_B(t, S; T, K, \sigma)$ is Black's call option formula with volatility σ, namely

$$c_B(t, S; T, K, \sigma) = S\mathcal{N}(d_+) - K(d_-), \qquad d_\pm = \frac{\ln\frac{S}{K} \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}$$

7.3.3 Spread options

We now sketch a crude approach for pricing spread options that is adequate for calibration purposes. A more accurate evaluation using copulas is presented later. Consider a spread option paying

$$V_{\text{spread}}(T) = (S_1(T) - S_2(T) - K)^+, \quad T \leq \min\{T_{j_1}, T_{j_2}\}$$

where $S_1(t) = S_{j_1,k_1-j_1}(t)$ and $S_2(t) = S_{j_2,k_2-j_2}(t)$ so that

$$\begin{aligned} dS_i(t) &= \mathcal{O}(dt) + \sqrt{z(t)}\varphi(S_i(t))\lambda_{S_i}(t)^\top dW^B(t), \\ \lambda_{S_i}(t) &\simeq \sum_{n=j}^{k_i-1} \omega_{S_i,n}(0)\lambda_n(t) \end{aligned}$$

Its time-0 value in the T-forward measure is given by

$$V_{\text{spread}}(0) = P(0,T)E^T\left[(S_1(T) - S_2(T) - K)^+\right]$$

Simplification. Assume that the spread $\epsilon(T) = S_1(T) - S_2(T)$ is a Gaussian with mean and variance

$$\begin{aligned} E^T[\epsilon(T)] &= E^T[S_1(T)] - E^T[S_2(T)] \simeq S_1(0) - S_2(0), \\ \text{Var}[\epsilon(T)] &= \sum_{i=1}^{2} \varphi(S_i(0))^2 z_0 \int_0^T \|\lambda_{S_1}(t)\|^2\, dt \\ &\quad -2\rho_{\text{term}}(0,T)z_0 \prod_{i=1}^{2} \varphi(S_i(0)) \left(\int_0^T \|\lambda_{S_1}(t)\|^2\, dt\right)^{1/2} \\ \rho_{\text{term}}(T,T') &= \text{Corr}(S_1(T) - S_1(T'), S_2(T) - S_2(T')) \\ &=\simeq \frac{\int_{T'}^T \lambda_{S_1}(t)^\top \cdot \lambda_{S_2}(t)}{\sqrt{\int_{T'}^T \|\lambda_{S_1}(t)\|^2\, dt} \cdot \sqrt{\int_{T'}^T \|\lambda_{S_2}(t)\|^2\, dt}} \end{aligned}$$

Bachelier's formula then yields

$$
\begin{aligned}
V_{\text{spread}}(0) &= P(0,T)\sqrt{\mathrm{Var}^T(\epsilon(T))}\cdot\left[d\Phi(d)+\phi(d)\right],\\
d &= \frac{E^T[\epsilon(T)]-K}{\sqrt{\mathrm{Var}^T(\epsilon(T))}}
\end{aligned}
$$

7.4 Calibration

Assume that a tenor structure $T_0 < T_1 < \cdots < T_N$ has been selected, that we have decided upon the number of Brownian motions m to be used and that we have selected the basic form of the LM model (e.g. local/stochastic volatility) to be deployed. To complete the model specification we need to establish the m-dimensional deterministic volatility vectors $\lambda_k(\cdot)$, $k = 1, \ldots, N-1$. We follow these steps:

(i) Prescribe the basic form of $\|\lambda_k(t)\|$ by introducing discrete time- and tenor-grids.

(ii) Use correlation information to obtain $\lambda_k(t)$ from $\|\lambda_k(t)\|$.

(iii) Choose a set of observable securities against which to calibrate the model and establish the norm to be used for calibration.

(iv) Recover $\lambda_k(t)$ by norm optimization.

7.4.1 Calibration objective function

Assume that our calibration targets are

$$
V_{swaption,1}, \ldots, V_{swaption,N_S} \quad V_{cap,1}, \ldots, V_{cap,N_C}
$$

Denote by $\hat{V}$ their quoted market prices and by $\bar{V}(G)$ their model-generated prices as functions of the volatility gird G. We will use the following calibration objective function:

$$
\begin{aligned}
\mathcal{I}(G) = \frac{w_S}{N_S}\sum_{i=1}^{N_S}\left(\bar{V}_{swaption,i}(G)-\hat{V}_{swaption,i}\right)^2 + \frac{w_C}{N_C}\sum_{i=1}^{N_C}\left(\bar{V}_{cap,i}(G)-\hat{V}_{cap,i}\right)^2 \\
\frac{w_{\partial t}}{N_x N_t}\sum_{i,j=1}^{N_t,N_x}\left(\frac{\partial G_{i,j}}{\partial t_i}\right)^2 + \frac{w_{\partial x}}{N_x N_t}\sum_{i,j=1}^{N_t,N_x}\left(\frac{\partial G_{i,j}}{\partial x_j}\right)^2 \\
\frac{w_{\partial t^2}}{N_x N_t}\sum_{i,j=1}^{N_t,N_x}\left(\frac{\partial^2 G_{i,j}}{\partial t_i^2}\right)^2 + \frac{w_{\partial x^2}}{N_x N_t}\sum_{i,j=1}^{N_t,N_x}\left(\frac{\partial^2 G_{i,j}}{\partial x_j^2}\right)^2
\end{aligned}
\tag{84}
$$

where $w_S, w_C, w_{\partial t}, w_{\partial x}, w_{\partial t^2}, w_{\partial x^2}$ are exogenously specified weights which determine the trade-off between accuracy and smoothness. Here is what all the terms above account for:

1. Terms (i) and (ii) measure how well the model is capable of reproducing the supplied market prices.
2. Term (iii) measures the degree of volatility term structure time homogeneity, and penalizes volatility functions that vary too much over calendar time.
3. Term (iv) measures the smoothness of the calendar time evolution of volatilities and penalizes deviations from linear evolution.
4. Term (v) and (vi) measure constancy and smoothness of in the tenor direction.

7.4.1.1 Error function for implied volatilities

The error (objective) function (84) is often applied implied volatilities as opposed to outright prices, so one normally institutes a pre-processing step where the market price of each swaption $\hat{V}_{swaption,i}$ is converted into a constant implied volatility $\hat{\lambda}_{S_i}$, in such a way that the scalar SDE for the swap rate S_i underlying the swaption $V_{swaption,i}$,

$$dS_i(t) = \sqrt{z(t)}\hat{\lambda}_{S_i}\varphi(S_i(t))dY_i(t)$$

reproduces the observed market price. If $\bar{\lambda}_{S_i}(G)$ denotes the corresponding model volatility (and $\bar{\lambda}_{C_i}(G)$ denotes the corresponding implied volatility for caps), one obtains an alternative calibration norm

$$\mathcal{I}(G) = \frac{w_S}{N_S}\sum_{i=1}^{N_S}\left(\bar{\lambda}_{S_i}(G) - \hat{\lambda}_{S_i}\right)^2 + \frac{w_C}{N_C}\sum_{i=1}^{N_C}\left(\bar{\lambda}_{C_i}(G) - \hat{\lambda}_{C_i}\right)^2 + \cdots$$

There are two main *reasons for proceeding in this fashion*:

- The relative scaling of individual swaptions and caps is more natural (high-valued trades tend to be overweighed using outright prices).
- In some models, the implied volatilities $\hat{\lambda}_{S_i}$ and $\hat{\lambda}_{C_i}$ can be computed by simple time integration, so using them for calibration avoids the need of using time-consuming option pricing formulas to obtain $\bar{V}_{swaption,i}$ and so forth. For instance, in Corollary 7.4, where we illustrated how to price swaptions under displaced log-normal Libor rates

  $$dL_n(t) = [bL_n(t) + (1-b)L_n(0)]\,\lambda_n(t)^\top dW^{n+1}(t), \quad n = 1,\dots,N-1$$

  we saw that the time-0 price of a T_j-maturity swaption is given by Black's call option formula

  $$V_{swaption,j}(0) = A(0)c_B\left(0, \frac{S(0)}{b}; T_j, c - S(0) + \frac{S(0)}{b}; b\bar{\lambda}_{S_j}\right)$$

  with term swap rate volatility given by

  $$\begin{aligned}\bar{\lambda}_{S_j} &= \frac{1}{T_j}\left(\int_0^{T_j} \|\lambda_S(t)\|^2\, dt\right)^{1/2},\\ \lambda_S(t) &= \sum_{n=j}^{k-1}\omega_n(0)\lambda_n(t)\end{aligned}$$

  where the weights $\omega_n(t)$ are given in (80). Given $\hat{V}_{swaption,j}$, the implied volatility in this case would be $\hat{\lambda}_{S_j} := b\bar{\lambda}_{S_j}$.

7.4.2 Calibration algorithm

Fix a number of factors m and fix a tenor structure $0 = T_0 < T_1 < \cdots < T_N$ and assume that all the parameters in our Libor model have been fixed exogenously[19] except for the m-dimensional vectors $\lambda_k(t)$, $k = 1,\dots,N-1$, which we now seek to calibrate. Define functions $h : \mathbb{R}^2 \to \mathbb{R}^m$ and $g : \mathbb{R}^2 \to \mathbb{R}$ as follows

$$\lambda_k(t) = h(t, T_k - t), \qquad \|\lambda_k(t)\| = g(t, T_k - t)$$

[19]Namely, assume that we have exogenously fixed the parameters defining the square-root process for $z(t)$ (i.e. θ, η) and the parameters defining the local volatility function $\varphi(\cdot)$ (i.e. b if we use $\varphi(L_n(t)) = bL_n(t)+(1-b)L_n(0)$.

Fix a time/tenor grid

$$\{t_i\} \times \{x_j\}, \qquad i = 1, \ldots, N_t, \quad j = 1, \ldots, N_x, \quad N_t, N_x \sim 10:15$$

subject to

$$t_1 + x_{N_x} \geq T_N, \quad \text{and} \quad t_i + x_j \in \{T_n\}$$

have been selected, together with the number of Brownian motions m, a correlation matrix R, the set of calibration swaptions and caps and the weights in the calibration norm $\mathcal{I}$.

Starting from some guess for the $N_t \times N_x$ matrix G, $G_{ij} = g(t_i, x_j)$ (so that for some n we have $g(t_i, \overbrace{T_n - t_i}^{x_j}) = \|\lambda_n(t_i)\|$), we run the following iterative algorithm:

1. Given G, use interpolation to obtain the full norm volatility grid $\|\lambda_{n,k}\|$ for all Libor indices $k = 1, \ldots, N-1$ and all expiry indices $n = 1, \ldots, k$. This is done by assuming that, for each k, the norm $\|\lambda_k(t)\|$ is piecewise constant in t

$$\begin{aligned} \|\lambda_k(t)\| &= \sum_{n=1}^{k} \mathbb{I}_{[T_{n-1}, T_n)}(t) \|\lambda_{n,k}\| \\ \|\lambda_{n,k}\| &= w_{++} G_{i,j} + w_{+-} G_{i,j-1} + w_{-+} G_{i-1,j} + w_{--} G_{i-1,j-1} \end{aligned}$$

and using linear interpolation in both dimensions of G.

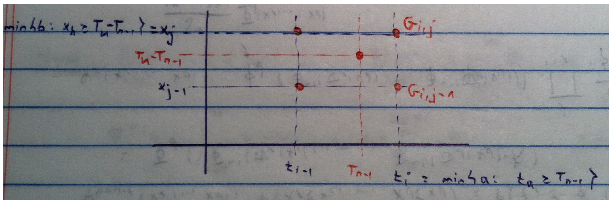

Figure 1: Computation of $\|\lambda_{n,k}\| = \|\lambda_k(t)\|_{|[T_{n-1},T_n)} = g(t, T_k - t)$

2. For each $n = 1, \ldots, N-1$, compute the $(N-n) \times m$ matrix $H(T_n)$,

$$[H(T_n)]_{ji} = h_i(T_n, T_{n+j-1} - T_n), \quad i = 1, \ldots, m, \quad j = 1, \ldots, N-n$$

as follows: by construction such $H(T_n)H(T_n)^\top$ will be the covariance matrix of the $(N-n)$-dimensional vector[20] $(L_n(T_n), \ldots, L_{N-1}(T_n))$. We can construct it by defining an instantaneous correlation $(N-n) \times (N-n)$ matrix (constructed from an estimated parametric form)

$$[R(T_n)]_{ij} = \mathrm{Corr}(L_i(T_n), L_j(T_n)), \quad i, j = n, \ldots, N-1 \tag{85}$$

and a diagonal volatility matrix

$$[c(T_n)]_{jj} = \|\lambda_{n,n+j-1}\|, \quad j = 1, \ldots, N-n$$

so that the instantaneous covariance matrix will be

$$C(T_n) = c(T_n) R(T_n) c(T_n)$$

[20]Indeed, the (i,j)-term of the matrix $H(T_n)H(T_n)^\top$ is

$$h(T_n, T_{n+i-1})^\top h(T_n, T_{n+j-1}) = \lambda_{n+i-1})(T_n)^\top \cdot \lambda_{n+j-1})(T_n) = \mathrm{Cov}(dL_{n+i-1}(T_n), dL_{n+j-1}(T_n))$$

We can finally solve for $H(T_n)$ in the equation

$$c(T_n)R(T_n)c(T_n) = H(T_n)H(T_n)^\top$$

via PCA.

(i) **Covariance PCA**: Let $\Lambda_m(T_n)$ be a diagonal matrix containing the m largest eigenvalues of $R(T_n)$, and let $e_m(T_n)$ be an $(N-n)\times m$ whose columns contain the corresponding eigenvectors. From the approximation[21]

$$c(T_n)R(T_n)c(T_n) \simeq e_m(T_n)\Lambda_m(T_n)e_m(T_n)^\top = e_m(T_n)\sqrt{\Lambda_m(T_n)}\sqrt{\Lambda_m(T_n)}e_m(T_n)^\top$$

it follows that

$$H(T_n) \simeq e_m(T_n)\sqrt{\Lambda_m(T_n)}$$

(ii) **Correlation PCA**: It is more efficient[22] to work with the correlation matrix directly and factor

$$R(T_n) = D(T_n)D(T_n)^\top$$

where $D(T_n)$ is an $(N-n)\times m$ matrix obtained using PCA analysis, namely

$$D = \arg\min_{D'}\{h(D';R) \text{ s.t. } v(D) = \mathbf{1}\} \overset{[*]}{=} \arg\min_{D'} h(D'; R + \mathrm{diag}(\mu)$$

where $h(D;R) = tr\left((R-DD^\top)(R-DD^\top)^\top\right)$, $v(D)$ denotes the vector of diagonal elements of $DD^\top$ and μ^* is the solution of the equation

$$v(D_\mu) = \mathbf{1}.$$

The passage from a constrained minimization problem to an unconstrained one in [$*$] is justified by [AP10-2, Proposition 14.3.2].

3. Given $\lambda_k(\cdot)$ for all $k = 1,\ldots,N-1$, use the pricing formulas in Propositions 7.1 and 7.3 to model prices for all swaptions and caps in the calibration set.
4. Establish the value of $\mathcal{I}(G)$ by direct computation.
5. Update G and repeat steps $1-4$ until $\mathcal{I}(G)$ is minimized, using for instance the downhill simplex method.

Remark 7.2. We will see how to calibrate the LM model to skews and smiles in section 7.7.3.

7.4.3 Correlation calibration to spread options

In the above calibration algorithm, the instantaneous correlation matrix R (85) was specified exogenously by calibrating a parametric form such as (77) to an empirical forward rate correlation matrix (76). Instead, one can attempt to imply R from market data.

Assume the matrix R is time-inhomogeneous and is specified as some parametric function of a low-dimension (unknown) parameter-vector[23] ξ

$$R = R(\xi)$$

For this, let $\mathcal{I}(G,\xi)$ be our old objective function (which depends on ξ, since the prices of caps and swaptions depend on the correlation matrix). Introduce:

[21] Which minimizes the norm $Tr\left[(C(T_n) - e_m(T_n)\Lambda_m(T_n)e_m(T_n)^\top)\cdot(C(T_n) - e_m(T_n)\Lambda_m(T_n)e_m(T_n)^\top)\right]$.

[22] Even though the PCA decomposition of a correlation matrix is technically more difficult, the correlation PCA is independent of the matrix $c(T_n)$ and as such it won't have to be updated when we update guesses for the G matrix in a calibration search loop.

[23] To be determined in the calibration procedure, along with the elements of G.

- Set of market-observable spread options $\hat{V}_{\text{spread},1}, \ldots, \hat{V}_{\text{spread},N_{SP}}$.
- Corresponding model-based prices $\bar{V}_{\text{spread},1}(G,\xi), \ldots, \bar{V}_{\text{spread},N_{SP}}(G,\xi)$, computed using ??

Update the old norm $\mathcal{I}$ to the norm

$$\mathcal{I}^*(G,\xi) = \mathcal{I}(G,\xi) + \frac{\omega_{NS}}{N_{SP}} \sum_{i=1}^{N_{SP}} \left(\bar{V}_{\text{spread},i}(G,\xi) - \hat{V}_{\text{spread},i}\right)^2$$

7.5 Monte Carlo simulation

Once the LM model has been calibrated to market data, we can proceed to use the parameterized model for the pricing and risk management of non-vanilla options. Given the large number of Markov state variables (the full number of Libor forward rates), one must nearly always rely on Monte Carlo simulation.

Goal: given a probability measure and the state of the Libor forward curve at time t, we need to move the entire Libor curve forward to time $t+\Delta$, $\Delta > 0$.

7.5.1 Euler-type schemes

Assume we stand at time t and we know $L_{q(t)}(t), L_{q(t)+1}(t), \ldots, L_{N-1}(t)$. We seek to advance to time $t+\Delta$ and construct a sample of $L_{q(t+\Delta)}(t+\Delta), L_{q(t+\Delta)+1}(t+\Delta), \ldots, L_{N-1}(t+\Delta)$. If $q(t+\Delta)$ exceeds $q(t)$, some of the front-end forward rates expire and drop off the curve as we move to $t+\Delta$. Working in the spot measure Q^B, the general LM model dynamics are

$$\begin{aligned} dL_n(t) &= \sigma_n(t)^\top \left[\mu_n(t)dt + dW^B(t)\right] \\ \mu_n(t) &= \sum_{j=q(t)}^{n} \frac{\tau_j \sigma_j(t)}{1+\tau_j L_j(t)} \end{aligned}$$

The Euler and log-Euler schemes for this model are:

$$\begin{aligned} \hat{L}_n(t+\Delta) &= \hat{L}_n(t) + \sigma_n(t)^\top \left[\mu_n(t)\Delta + \sqrt{\Delta}Z\right], \\ \hat{L}_n(t+\Delta) &= \hat{L}_n(\Delta) \exp\left[\frac{1}{\hat{L}_n(t)} \sigma_n(t)^\top \left(\left(\mu_n(t) - \frac{1}{2}\frac{\sigma_n(t)}{\hat{L}_n(t)}\right)\Delta + \sqrt{\Delta}Z\right)\right] \end{aligned}$$

where $Z \sim N = (0, I_m)$ is a vector of m independent Gaussian draws.

7.5.2 Special-purpose schemes with drift Predictor-Corrector

In integrated form, the general LM dynamics become

$$\begin{aligned} L_n(t+\Delta) &= L_n(t) + \int_t^{t+\Delta} \sigma_n(u)^\top \mu_n(u)\, du + \int_t^{t+\Delta} \sigma_n(u)^\top dW^B(u) \\ &\overset{not}{=} L_n(t) + D_n(t,t+\Delta) + M_n(t,t+\Delta) \end{aligned} \tag{86}$$

On occasions high-performance special-purpose schemes exist for simulation of $M_n(t,t+\Delta)$. Generating $\hat{M}(t,t+\Delta)$ by a special-purpose scheme, we can use a predictor-corrector scheme for L_n:

$$\begin{aligned}
\bar{L}_n(t+\Delta) &= \hat{L}_n(t) + \sigma_n(t,\hat{\mathbf{L}}(t))^\top \mu_n(t,\hat{\mathbf{L}}(t))\Delta + \hat{M}(t,t+\Delta) \\
\hat{L}_n(t+\Delta) &= \hat{L}_n(t) + \theta_{PC}\sigma_n(t,\hat{\mathbf{L}}(t))^\top \mu_n(t,\hat{\mathbf{L}}(t))\Delta \\
&\quad +(1-\theta_{PC})\sigma_n(t+\Delta,\hat{\mathbf{L}}(t+\Delta))^\top \mu_n(t+\Delta,\hat{\mathbf{L}}(t+\Delta))\Delta \\
&\quad +\hat{M}(t,t+\Delta) \qquad (87) \\
\hat{M}(t,t+\Delta) &= \sigma_n(t)^\top \sqrt{\Delta} Z
\end{aligned}$$

where $\mathbf{L}(t) = (L_1(t),\ldots,L_{N-1}(t))^\top$, $\hat{\mathbf{L}}$ is the semi-implicit corrector scheme, $\bar{\mathbf{L}}$ is the explicit predictor scheme and $\theta_{PC} \in [0,1]$ is the parameter that determined the amount of implicitness.

7.5.2.1 Lagging Predictor-Corrector scheme

7.6 Interpolation of front / back stubs

So far, we have seen how to obtain, at any time t, a vector of forward Libor rates

$$\mathbf{L}(t) = (L_{q(t)}(t), L_{q(t)+1}, \ldots, L_{N-1}(t))$$

on a pre-specified tenor structure $0 \le T_0 < \cdots < T_N$.

In order to be able to compute $P(t,T_n)$ for all n we need to additionally establish the*front stub* factor $P(t,T_{q(t)})$. Indeed, recall that

$$P(t,T_n) = P(t,T_{q(t)}) \prod_{i=q(t)}^{n-1} \frac{1}{1+\tau_i L_i(t)}$$

Also, in order to compute $P(t,T)$ for arbitrary T, the *back stub* discount factor $P(t,T,T_{q(T)}) = \frac{P(t,T_{q(T)})}{P(t,T)}$ is needed since

$$P(t,T) = \frac{P(t,T_{q(T)})}{P(t,T,T_{q(T)})} = \frac{1}{P(t,T,T_{q(T)})} \overbrace{P(t,T_{q(t)}) \prod_{i=q(t)}^{q(T)-1} \frac{1}{1+\tau_i L_i(t)}}^{P(t,T_{q(T)})}$$

The front and back stubs cannot be generally expressed in terms if Libor rates.

7.6.1 Back stub interpolation

A simple idea to obtain $P(t,T,T_m)$ is to interpolate between $P(t,T_m,T_m) = 1$ and $P(t,T_{m-1},T_m) = \frac{1}{1+\tau_{m-1}L_{m-1}(t)}$, for instance setting

$$P(t,T,T_m) = \frac{T-T_{m-1}}{T_m-T_{m-1}} + \frac{T_m-T}{T_m-T_{m-1}} P(t,T_{m-1},T_m)$$

Interpolations like these violate the basic no-arbitrage condition

$$P(0,T_{m-1},T_m) = E^{T_{m-1}}\left[P(t,T_{m-1},T_m\right]$$

One way to remedy this is to consider choosing na arbitrary constant $\alpha(T)$ and setting

$$P(t,T_{m-1},T_m) = P(0,T_{m-1},T_m) + \alpha(T)\left(P(t,T_{m-1},T_m) - P(0,T_{m-1},T_m)\right)$$

subject to the consistency conditions $\alpha(T_m) = 1$ and $\alpha(T_{m-1}) = 0$. Note

$$
\begin{aligned}
dP(t, T_{m-1}, T) &= \mathcal{O}(dt) + \alpha(T) dP(t, T_{m-1}, T_m), \\
\frac{dP(t, T_{m-1}, T)}{P(t, T_{m-1}, T)} &= \mathcal{O}(dt) + [\sigma_P(t, T_{m-1}) - \sigma_P(t, T)]\, dW(t), \\
\frac{dP(t, T_{m-1}, T_m)}{P(t, T_{m-1}, T_m)} &= \mathcal{O}(dt) + [\sigma_P(t, T_{m-1}) - \sigma_P(t, T_m)]\, dW(t)
\end{aligned}
$$

We may thus approximate $\alpha(T)$ as

$$
\alpha(T) \simeq \frac{P(0, T_{m-1}, T)}{P(0, T_{m-1}, T_m)} \frac{\|\sigma_P(t, T_{m-1}) - \sigma_P(t, T)\|}{\|\sigma_P(t, T_{m-1}) - \sigma_P(t, T_m)\|}
$$

This approximation turns the problem of interpolating bond prices to that of interpolating bond volatilities. For instance, in the one-dimensional Gaussian model with constant mean reversion κ we have

$$
\sigma_P(t, T) = \sigma(t) \frac{1 - e^{-\kappa(T-t)}}{\kappa}
$$

Plugging this into the approximation above yields

$$
\alpha(T) \simeq \frac{P(0, T_{m-1}, T)}{P(0, T_{m-1}, T_m)} \frac{1 - e^{-\kappa(T - T_{m-1})}}{1 - e^{-\kappa(T_m - T_{m-1})}}
$$

where κ could be either set as part of the user input or obtained by best-fitting the Gaussian parametric form to the volatility structure of the LM model.

7.6.2 Back stub interpolation inspired by the Gaussian model

From the bond reconstitution formula in the Gaussian model we have

$$
\begin{aligned}
P(t, T_{m-1}, T_m) = P(0, T_{m-1}, T_m) \times \exp\left[-(G(T_{m-1}, T_m) e^{-\kappa(T_{m-1}-t)} x(t')\right] \\
\times \exp\left[-\frac{1}{2}(G(t, T_m)^2 - G(t, T_{m-1})^2) y(t)\right]
\end{aligned}
$$

where

$$
G(t, T) = \frac{1 - e^{-\kappa(T-t)}}{\kappa}, \qquad y(t) = e^{-2\kappa t} \int_0^t e^{2\kappa s} \sigma(s)^2\, ds
$$

A simple computation yields

$$
\begin{aligned}
P(t, T_{m-1}, T) &= P(0, T_{m-1}, T) \left(\frac{P(t, T_{m-1}, T_m)}{P(0, T_{m-1}, T_m)} \right)^{\frac{G(T_{m-1}, T)}{G(T_{m-1}, T_m)}} \\
&\quad \times \exp\left[\frac{1}{2} \frac{G(T_{m-1}, T)}{G(T_{m-1}, T_m)} (G(t, T_m)^2 - G(t, T_{m-1})^2) y(t)\right] \\
&\quad \times \exp\left[-\frac{1}{2}(G(t, T)^2 - G(t, T_{m-1})^2) y(t)\right]
\end{aligned}
$$

where the volatility $\sigma(t)$ and the mean reversion parameter κ can be obtained by fitting the Gaussian volatility structure to the volatilities of Libor rates generated by the LM model itself.

7.6.3 Front stub interpolation inspired by the Gaussian model

Using the forward bond prices we obtained in the previous subsection $P(t', T_m, T_{m+1})$, $t' \geq m = q(t')$, we can now construct $P(t', T_m)$ from these by employing a bond reconstitution formula from a Gaussian model.

Assume that for short tenors, the LM model can be locally approximated by a one-factor Gaussian model with constant mean reversion. In particular we have

$$\begin{aligned}
P(t', T_m) &= P(0, t', T_m) \exp\left[-G(t', T_m)x(t') - \frac{1}{2}G(t', T_m)^2 y(t')\right], \\
P(t', T_m, T_{m+1}) &= P(0, T_m, T_{m+1}) \times \exp\left[-(G(t', T_{m+1}) - G(t', T_m))x(t')\right] \\
&\quad \times \exp\left[-\frac{1}{2}(G(t', T_{m+1})^2 - G(t', T_m)^2)y(t')\right]
\end{aligned}$$

whereby one obtains

$$P(t', T_m) = P(0, t', T_m) \left(\frac{P(t', T_m, T_{m+1})}{P(0, T_m, T_{m+1})}\right)^{\frac{G(t', T_m)}{G(t', T_{m+1}) - G(t', T_m)}} \times \exp\left(\frac{1}{2}G(t', T_m)G(t', T_{m+1})y(t')\right)$$

where

- The short rate volatility $\sigma(t)$ can be approximated by the volatilities of the front Libor rate.
- The mean reversion parameter κ can be used to set the stub bond volatility near to its market-implied value.

7.7 Advanced approximation for swaption pricing via Markovian projection

So far we have seen how to calibrate the SV-LM model (74)-(75)

$$\begin{aligned}
dL_n(t) &= \sqrt{z(t)}\varphi(L_n(t))\lambda_n(t)^\top \left(\sqrt{z(t)}\mu_n(t)dt + dW^B(t)\right) \\
\mu_n(t) &= \sum_{j=q(t)}^{n} \frac{\tau_j \varphi(L_j(t))\lambda_j(t)}{1 + \tau_j L_j(t)} \\
dz(t) &= \theta[z_0 - z(t)]dt + \eta\psi(z(t))dZ(t), \quad z(0) = 0
\end{aligned}$$

and how to use it to price non-vanilla derivatives via Monte Carlo discretization. In order to calibrate the model to liquid caps and swaptions, we developed fast pricing formulas (78)-(81) for the prices thereof, which in the case of swaptions required the use of approximated dynamics for the swap rate $S(t)$.

- When approximating the dynamics of the swap rate $S(t)$ to price swaptions via (82)

  $$dS(t) \simeq \sqrt{z(t)}\varphi(S(t))\|\lambda_S(t)\|dY^A(t)$$

  the skew of the swap rate is taken to be the same as the (common) skew of all Libor rates, yet numerical simulation shows that this is not the case.
- The model (74)

  $$dL_n(t) = \sqrt{z(t)}\varphi(L_n(t))\lambda_n(t)^\top \left(\sqrt{z(t)}\mu_n(t)dt + dW^B(t)\right)$$

  uses the same time-homogeneous local volatility function $\varphi(\cdot)$ for all Libor rates, which implies that swaptions of all tenors and expiries have essentially the same volatility skew, a model feature which is again inconsistent with market reality.

We thus consider the following generalization the old LM model specification (74)-(75)

$$\begin{aligned} dL_n(t) &= \sqrt{z(t)}\varphi_n(t, L_n(t))\lambda_n(t)^\top dW^{T_{n+1}}(t), \quad n = 1, \ldots, N-1 && (88) \\ dz(t) &= \theta[z_0 - z(t)]dt + \eta(t)\psi(z(t))dZ(t), \quad z(0) = 0 && (89) \end{aligned}$$

and assume

$$\begin{aligned} \varphi_n(t, L_n(0)) &= 1 \\ \varphi_n(t, x) &\simeq 1 + b_n(t)\left[x - L_n(0)\right], \\ b_n(t) &= \partial_x \varphi_n(t, L_n(0)) \end{aligned}$$

The dynamics of the swap rate $S(t) = S_{j,k-k}(t)$ in this model are

$$\begin{aligned} dS(t) &= \sqrt{z(t)}\lambda_S(t, \mathbf{L}(t))^\top dW^A(t) && (90) \\ \lambda_S(t, \mathbf{L}(t)) &= \sum_{n=j}^{k-1} \frac{\partial S(t)}{\partial L_n(t)} \varphi_n(t, L_n(t))\lambda_n(t) \end{aligned}$$

or in one-dimensional form

$$dS(t) = \sqrt{z(t)}\|\lambda_S(t, \mathbf{L}(t))\| dY^A(t)$$

Using standard results on Markovian projection, we can derive the following approximation.

Proposition 7.5 (Swap rate dynamics approximation)**.** *For the purpose of European swaption valuation, the dynamics of the swap rate $S(t)$ in measure Q^A are approximately given by the following displaced log-normal stochastic volatility SDE*

$$\begin{aligned} dS(t) &\simeq \sqrt{z(t)}p_S(t)\left[1 + b_S(t)(S(t) - S(0))\right] dY^A(t) && (91) \\ p_S(t) &= \|\lambda_S\left(t, E^A[\boldsymbol{L}(t)]\right)\| && (92) \\ b_S(t) &= \frac{1}{p_S(t)} \sum_{n=j}^{k-1} \frac{\partial \|\lambda_S\left(t, E^A[\boldsymbol{L}(t)]\right)\|}{\partial L_n(t)} \frac{\int_0^t \lambda_n(u)^\top \lambda_S\left(u, \boldsymbol{L}(0)\right) du}{\int_0^t \|\lambda_S\left(u, \boldsymbol{L}(0)\right)\|^2 du} && (93) \end{aligned}$$

As shown in Proposition 7.3, the price of a T_j-maturity European swaption is given by

$$V_{swaption}(0) = A(0)E^A\left[(S(T_j) - c)^+\right]$$

and $S(T)$ is driven by the stochastic volatility model with time-dependent parameters

$$\begin{aligned} dS(t) &\simeq \sqrt{z(t)}p_S(t)\left[1 + b_S(t)(S(t) - S(0))\right] dY^A(t), \\ dz(t) &= \theta[z_0 - z(t)]dt + \eta\psi(z(t))dZ(t) \end{aligned}$$

This is a stochastic volatility model with time-dependent parameters, so using the time-averaging techniques of [AP10-1, Section 9.3] we have devised a fast way of pricing European swaptions.

The improvements over our early approximation essentially stem from the evaluation of the volatility function of the swap rate $\lambda_S(t, \boldsymbol{L}(t))$ at $\boldsymbol{L}(t) = E^A[\boldsymbol{L}(t)]$ rather than at $\boldsymbol{L}(t) = \boldsymbol{L}(0)$.

7.7.1 Advanced formula for the swap rate volatility

In order to compute the swap rate volatility (92)

$$p_S(t) = \|\lambda_S\left(t, E^A[\mathbf{L}(t)]\right)\|$$

we need to approximate the expectations $E^A[L_n(t)]$

Proposition 7.6 (Advanced formula for the swap rate volatility)**.** *For $j \leq n \leq k.1$, the expected value of the n-th Libor rate in the annuity measure is approximately given by*

$$\begin{aligned}
E^A[L_n(t)] &= L_n(0)[1 + c_n(t)] \\
c_n(t) &= \frac{1}{L_n(0)Q_n(0)} \sum_{i=j}^{k-1} \frac{\partial Q_n(0)}{\partial L_i(0)} \int_0^t \lambda_i^\top(s)\lambda_n(s)\,ds, \\
Q_n(t) &= \frac{A(t)}{P(t, T_{n+1})}
\end{aligned}$$

7.7.2 Advanced formula for the swap rate skew

We now show how to compute the swap rate skew (93).

Proposition 7.7 (Advanced formula for the swap rate skew)**.** *The time-dependent swaption skew $b_S(t)$ 93 is approximately given by*

$$\begin{aligned}
b_S(t) &= \sum_{i,n=j}^{k-1} \frac{r_{S,n}(t)}{r_{S,i}(t)} v_{i,n}(0)\xi_i(t) + \sum_{i=j}^{k-1} b_i(t)v_i(0)\xi_i(t) \\
r_{S,i}(t) &= \lambda_i(t)^\top \lambda_S(t, \boldsymbol{L}(0)), \\
r_S(t) &= \|\lambda_S(t, \boldsymbol{L}(0))\|^2 \\
\xi_i(t) &= \frac{r_{S,i}(t)\int_0^t r_{S,i}(u)\,du}{r_S(t)\int_0^t r_S(u)\,du}
\end{aligned}$$

7.7.3 Skew and smile calibration

The general model

$$\begin{aligned}
dL_n(t) &= \sqrt{z(t)}\varphi_n(t, L_n(t))\lambda_n(t)^\top dW^{T_{n+1}}(t), \quad n = 1, \ldots, N-1 \\
dz(t) &= \theta[z_0 - z(t)]dt + \eta(t)\psi(z(t))dZ(t), \quad z(0) = 0 \\
\varphi_n(t, x) &= 1 + b_n(t)\left[x - L_n(0)\right]
\end{aligned}$$

has enough flexibility to match (in addition to the correlation structure of Libor rates):

- ATM volatilities of all European swaptions, using $\|\lambda_S(t)\|$.
- Slopes of volatility smiles (skews), using $b_n(t)$.
- Curvatures of smiles for swaptions of different expiries and fixed tenor, using $\eta(t)$.

Consider N_S swaptions $\mathrm{V}_{\mathrm{swaption},1}, \ldots, \mathrm{V}_{\mathrm{swaption},\mathrm{N_S}}$ as calibration targets, where each $\mathrm{V}_{\mathrm{swaption,i}}$ represents a collection of swaptions with fixed tenor and different strikes.

We first convert the observed market prices $\hat{\mathrm{V}}_{swaption,1}, \ldots, \hat{\mathrm{V}}_{swaption,N_S}$ into market-implied SV parameters, namely implied volatilities $\hat{\lambda}_{S_i}$, skews $\hat{b}_{S_i}$ and volatilities of variance

$\hat{\eta}_{S_i}$ for $i = 1, \ldots, N_S$ (these conversions are routinely maintained and updated by trading desks, c.f. [AP10-1, Page 378]), presumably by a simple root-search[24].

As in the standard Libor model, consider a grid of calendar times $\{t_i\}_{i=1}^{N_t}$ and tenors $\{x_j\}_{j=1}^{N_x}$ and define:

- $[G_\lambda]_{i,j} = \|\lambda_n(t_i)\|$, where we recall that n is such that $T_n - t_i = x_j$.
- $[G_b]_{i,j} = b_n(t_i)$.
- $[G_\eta]_i = \eta(t_i)$.

The formulas from the previous section allow us to compute model term volatilities, skews and volatilities of variance by

$$\bar{\lambda}_{S_i}(G_\lambda, G_b, G_\eta), \quad \bar{b}_{S_i}(G_\lambda, G_b, G_\eta), \quad \bar{\eta}_{S_i}(G_\lambda, G_b, G_\eta)$$

Instead of defining a uniform norm incorporating all these terms (which would result in a hard multi-dimensional non-linear optimization problem), we define three norms[25]:

$$\begin{aligned}
\mathcal{I}_\lambda(G_\lambda, G_b, G_\eta) &= \frac{\omega_{S,\lambda}}{N_S} \sum_{i=1}^{N_S} \left(\bar{\lambda}_{S_i}(G_\lambda, G_b, G_\eta) - \hat{\lambda}_{S_i}\right)^2 + \cdots \\
\mathcal{I}_b(G_\lambda, G_b, G_\eta) &= \frac{\omega_{S,b}}{N_S} \sum_{i=1}^{N_S} \left(\bar{b}_{S_i}(G_\lambda, G_b, G_\eta) - \hat{b}_{S_i}\right)^2 + \cdots \\
\mathcal{I}_\eta(G_\lambda, G_b, G_\eta) &= \frac{\omega_{S,\eta}}{N_S} \sum_{i=1}^{N_S} (\bar{\eta}_{S_i}(G_\lambda, G_b, G_\eta) - \hat{\eta}_{S_i})^2 + \cdots
\end{aligned}$$

This leads to the following calibration algorithm:

1. Make initial guesses $\mathcal{I}_b^0$ and $\mathcal{I}_\eta^0$. For instance take

$$\mathcal{I}_b^0 = \frac{1}{N_S} \sum_{k=1}^{N_S} \hat{b}_{S_k}, \qquad \mathcal{I}_\eta^0 = \frac{1}{N_S} \sum_{k=1}^{N_S} \hat{\eta}_{S_k}$$

2. Use the calibration algorithm in section 7.4.2 to obtain

$$G_\lambda^1 := \arg\min_{G_\lambda} \mathcal{I}_\lambda(G_\lambda, G_b^0, G_\eta^0).$$

3. Iterate over G_b to obtain

$$G_b^1 := \arg\min_{G_b} \mathcal{I}_b(G_\lambda^1, G_b, G_\eta^0).$$

4. Iterate over G_η to obtain

$$G_\eta^1 := \arg\min_{G_\eta} \mathcal{I}_\eta(G_\lambda^1, G_b^1, G_\eta).$$

One could then iterate further starting with (G_b^1, G_η^1) on the first step.

[24] See the end of section 7.4.1 for a concrete example.

[25] Impact of changes in the Libor skews on term swaption volatilities, or their volatilities of variance, is rather small.

7.8 Other extensions

7.8.1 Evolving separate discount and forward rate curves

A single yield curve for discounting and calculating Libor rates us not always compatible with no-arbitrage constraints of cross-currency markets. In general, separating the discounting curve from the forward curve will ensure that linear instruments (i.e. swaps and bonds) are correctly priced at time 0.

7.8.2 SV models with non-zero correlation

So far we have assumed that the scalar Brownian motion $Z(t)$ driving the variance process was independent of all the components of the m-dimensional Brownian motion $W(t)$. We can relax this assumption and assume a non-zero deterministic correlation vector $\rho(t)$ between $Z(t)$ and $W^B(t)$.

7.8.3 Multi-stochastic volatility extensions

In the specification (74)-(75) of the LM model, a single stochastic volatility process $\sqrt{z(t)}$ is used to scale the diffusion coefficients of all forward rates. While sufficiently rich to introduce the volatility smile for all European swaptions, it presents some limitations.

(i) Interest rate exotics link to the spread of two CMS rates derive their values from the distribution of the *slope* of the interest rate curve and a common stochastic volatility factor applied to all rates does not always allow for sufficient control over the distribution of the slope of the interest rate curve.

(ii) The are of multi-dimensional stochastic volatility interest rate modeling is quite new.

8 Single-rate vanilla derivatives

Here we study derivatives whose payoffs can be decomposed as a linear combination of functions of individual rate observations (either exactly or approximately).

8.1 European swaptions

Given a tenor structure $0 < T_0 < \cdots < T_N$, consider swaptions of different expiries $\{T_n\}_{n=0,\ldots,N-1}$ that can be exercised into swaps that start at T_n and cover m periods. Define

$$A_{n,m}(t) = \sum_{i=n}^{n+m-1} \tau_i P(t, T_{i+1}), \qquad S_{n,m}(t) = \frac{1}{A(t)} \sum_{i=n}^{n+m-1} \tau_i P(t, T_{i+1}) L_i(t)$$

The time-t value of the m-period swaption expiring at T_n is

$$V_{n,m}(t) = A_{n,m}(t) E_t^{n,m} \left[(S_{n,m}(t) - k)^+ \right]$$

For each swap rate $S_{n,m}(t)$ we specify stochastic volatility (SV) dynamics in the annuity measure

$$\begin{aligned} dS_{n,m}(t) &= \lambda_{n,m} \left[b_{n,m} S_{n,m}(t) + (1 - b_{n,m}) S_{n,m}(0) \right] \sqrt{z_{n,m}(t)} dW^{n,m}(t) \\ dz_{n,m}(t) &= \theta \left[1 - z_{n,m}(t) \right] dt + \eta_{n,m} \sqrt{z_{n,m}} dZ^{n,m}(t) \end{aligned} \tag{94}$$

where recall

(i) θ is the mean-reversion of variance

- It controls the speed at which deviations of $z(\cdot)$ away from z_0 are pulled back towards this level. Increasing θ decreases the long-term variance of $z(\dot{})$.
- We assume it is global: same parameter for all swaptions.
- The idea is to choose θ in order to minimize the variability of $\{\eta_{n,m}\}_{n,m}$ across different expiries n.

(ii) $\eta_{n,m}$ is the volatility of variance and controls the curvature of the volatility smile.

Swaption calibration is performed individually for each grid point: fix a pair maturity-tenor (n, m). Suppose a collection of strikes $K_1, \ldots, K_J$, along with corresponding market prices $\hat{V}_1, \ldots, \hat{V}_J$ (where we omit n, m from the notation since they are now fixed). The optimal parameters λ^*, b^*, η^* are obtained as the solution of the following minimization problem

$$(\lambda^*, b^*, \eta^*) = \text{argmin}_{\{\lambda,b,\eta\}} \mathcal{I}(\lambda, b, \eta), \quad \mathcal{I}(\lambda, b, \eta) = \sum_{j=1}^{J} w_J \left(V(K_j; \lambda, b, \eta) - \hat{V}_j \right)^2 \tag{95}$$

The algorithm to calibrate a SV-model to swaption prices is hence the following:

1. Choose θ.
2. For each maturity-tenor pair (n, m), find the optimal parameters $(\lambda^*_{n,m}, b^*_{n,m}, \eta^*_{n,m})$ by solving the optimization problem (95).
3. If the $\eta_{n,m}$'s increase (decrease) with n, reduce (increment) θ and iterate.

8.2 Caps and floors

8.3 Terminal swap rate (TSR) models

- Pricing swaptions and caps is simple because valuation requires only knowledge of the terminal distribution of a single swap rate in the appropriate annuity measure.
  1. Given a tenor structure $0 < T_0 \leq \cdots \leq T_N$, consider swaptions of different expiries $\{T_n\}_{n=0,\ldots,N-1}$ that can be exercised into swaps that start at T_n and cover m periods. Define

     $$A_{n,m}(t) = \sum_{i=n}^{n+m-1} \tau_i P(t, T_{i+1}), \qquad S_{n,m}(t) = \frac{1}{A(t)} \sum_{i=n}^{n+m-1} \tau_i P(t, T_{i+1}) L_i(t)$$

     The time-t value of the m-period swaption expiring at T_n is

     $$A_{n,m}(t) E_t^{n,m} \left[(S_{n,m}(t) - k)^+ \right]$$

  2. A cap...
- Much more common a relatively simple payoffs that depend on the rate $S(T)$, but which also require the knowledge of additional discount bonds.
- **Idea underlying TSR models**: instead of using a full-blown term structure model, it's better to functionally link the values of discount bonds on the date T to the driving rate $S(T)$ through certain approximations.

Given a tenor structure $0 < T_0 < T_1 < \cdots < T_N$ and defining the annuity and swap rate as usual

$$A(t) = A_{0,N}(t) = \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}), \qquad S(t) = S_{0,N}(t) = \frac{P(t,T) - P(t,T_N)}{A(t)}$$

the value of a derivative with payoff X in the annuity measure is

$$V(0) = A(0) E^A \left[\frac{X}{A(T)} \right] \tag{96}$$

If $\{P(T,M)\}_{M \geq T}$ is a family of discount bonds of various maturities (observed at time T), a **TSR model exogenously specifies a family of maturity-indexed functions** $\{\pi(\cdot, M)\}_{M \geq T}$ such that

$$P(T,M) = \pi(S(T), M), \quad \forall M \geq T \tag{97}$$

satisfying certain conditions

(i) **No-arbitrage condition**: when applying the valuation formula (96) to the specification (97) one should be able to reproduce the initial discount bond prices

$$P(0,M) = A(0) E^A \left[\frac{\pi(S(T),M)}{A(T)} \right] = A(0) E^A \left[\frac{\pi(S(T),M)}{\sum_{n=0}^{N-1} \tau_n \pi(S(T), T_{n+1})} \right], \quad M \geq T \tag{98}$$

(ii) **Consistency condition**: the swap rate $S(T)$ itself is a function of discount factors $S(T) = \frac{1-P(T,T_N)}{\sum_{n=0}^{N-1} \tau_n P(T,T_{n+1})}$ so we need to impose that

$$x = \frac{1 - \pi(x, T_N)}{\sum_{n=0}^{N-1} \tau_n \pi(S(T), T_{n+1})}, \qquad \forall x \tag{99}$$

(iii) **Reasonable family of functions**: we generally require the functions $\{\pi(\cdot, M)\}_M$ to range between 0 and 1, to be monotonic in M and to be continuous in (x, M). Generally one chooses a parametric class for the functions $\{\pi(\cdot, M)\}_M$ and then one chooses functions within the parametric class satisfying the no-arbitrage and consistency conditions.

8.3.1 Linear TSR models

We use the linear specification

$$\frac{\pi(x, M)}{\sum_{n=0}^{N-1} \tau_n \pi(x, T_{n+1})} = a(M)x + b(M), \quad M \geq T \tag{100}$$

for deterministic functions $a(\cdot), b(\cdot)$.

- The no-arbitrage condition implies that

$$b(M) = \frac{P(0, M)}{A(0)} - a(M)S(0) \tag{101}$$

- The consistency condition implies that

$$b(T_0) = b(T_N) \tag{102}$$
$$a(T_0) = 1 + a(T_N) \tag{103}$$

- The specification (100) itself imposes

$$\sum_{n=0}^{N-1} \tau_n a(T_{n+1}) = 0, \tag{104}$$
$$\sum_{n=0}^{N-1} \tau_n b(T_{n+1}) = 1. \tag{105}$$

To complete the model specification one may proceed as follows:

(i) Choose coefficients $\{a(T_1), \ldots, a(T_N)\}$ subject to condition (104), as described below.

(ii) Calculate $a(T) = a(T_0) = 1 + a(T_N)$ from condition (103).

(iii) Calculate the rest of $a(M)$'s by interpolation of $\{a(T_1), \ldots, a(T_N)\}$.

(iv) Calculate the $b(M)$'s by condition (101).

The coefficients $\{a(T_1), \ldots, a(T_N)\}$ can be connected to a mean reversion parameter: this will not only reduce the number of parameters required, will also parameterize the model with a single parameter that has a strong financial interpretation. The argument on pages 713-714 on [AP10-2] shows that one may choose

$$\begin{aligned} a(M) &= \frac{P(0, M)\,(\gamma - G(T, M))}{P(0, T_N)G(T, T_N) + S(0)\gamma} \\ \gamma &= \frac{\sum \tau_n P(0, T_{n+1}) G(T, T_{n+1})}{\sum \tau_n P(0, T_{n+1})} \\ G(t, T) &= \int_t^T e^{-\int_t^u \kappa(s)\, ds}\, du \end{aligned}$$

This also eliminates the need to interpolate in (iii) and facilitates better risk management (c.f. discussion on page 714 in [AP10-3]).

8.4 Libor-in-Arrears

A Libor-in-Arrears cash flow pays the Libor rate on the date when it fixes, rather than on the date it matures. We focus on a single cash flow, the valuation of a full strip following by additivity.Let $T > 0$ be the start date, and M the end date of the period covered by Libor rate $L(t) \equiv L(t; T, M) = \frac{P(t,T) - P(t,M)}{\tau P(t,M)}$, with $\tau = M - T$. The time-0 value of a Libor-in-Arrears cash flow is

$$V_{LIA}(0) = \beta(0) E\left[\frac{L(T)}{\beta(T)}\right]$$

One would then naturally switch to the T-forward measure, but traded caplets provide information about the distribution of the Libor rate in the M-forward measure, not the T-forward measure[26].

The valuation formula in the M-forward measure is

$$\begin{aligned} V_{LIA}(0) &= P(0, M) E^M\left[\frac{L(T)}{P(T, M)}\right] = P(0, M) E^M\left[(1 + \tau L(T)) P(T, M)\right] \\ &= P(0, M)\left(L(0) + \tau E^M\left[L^2(T)\right]\right) \end{aligned}$$

The term $E^M\left[L^2(T)\right]$ can be computed by Proposition **??**

$$E^M\left[L^2(T)\right] = L^2(0) + 2\int_{-\infty}^{L(0)} p(0, L(0); T, K)\, dK + 2\int_{L(0)}^{\infty} c(0, L(0); T, K)\, dK \tag{106}$$

where $p(t, L; T, K)$ and $c(t, L; T, K)$ are undiscounted values of put and call options on the rate $L(T)$ with strike K (i.e. simple undiscounted floorlets and caplets).

8.5 Libor-with-Delay

A Libor-with-delay cash-flow pays the Libor rate $L(t, T, M)$ at some arbitrary payment time $T_p \geq T$. In the M-forward measure we have

$$V_{LD}(0) = P(0, M) E^M\left[\frac{P(T, T_p)}{P(T, M)} L(T)\right]$$

and we can now use a TSR model to express $P(T, T_p)$ in terms of the Libor rate.

Alternatively one can draw inspiration from the quasi-Gaussian model to express $P(T, T_p)$ in terms of $L(T)$. Recall that

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

Approximating $y(T)$ by a deterministic function $\bar{y}(T) := E[y(T)]$ and denoting by $X(T, \ell)$ the inverse function in x to $L(T, x, \bar{y}(T))$ we can write

$$P(T, T_p) = P(T, T_p; X(T, L(T)), \bar{y}(T)) \tag{107}$$

[26] Note that $L(t) = L(t; T, M)$ is a martingale in the M-forward measure, but not in the T-forward measure, so $E^T[L(T)] \neq L(0)$. To characterize this situation one defines the *Libor-in-Arrears convexity adjustment*

$$D_{LIA}(0) \stackrel{def}{=} E^T[L(T)] - L(0).$$

In general, by *convexity* we generally mean the difference of valuations under different measures (in this case the difference between the measure appropriate for the given payment date and the measure in which the market rate is a martingale).

Assuming a linear relationship between the Libor rate and the state variable x

$$L(T) \simeq L(T,0,0) + \frac{\partial L}{\partial x}(T,0,0)x(T) \Longrightarrow \mathrm{Var}(L(T)) \simeq \left(\frac{\partial L}{\partial x}(T,0,0)\right)^2 \mathrm{Var}(x(T))$$

whence

$$\bar{y}(T) = \mathrm{Var}(x(T)) \simeq \left(\frac{\partial L}{\partial x}(T,0,0)\right)^{-2} \mathrm{Var}(L(T))$$

Thus given $P(T,T_p)$, we can solve equation (107) implicitly for $L(T)$ to obtain our sought for relation. Note that this procedure might result in a mild violation of the no-arbitrage condition.

8.6 CMS and CMS-Linked cash flows

A CMS (linked) cash-flow pays (a function of) the swap rate $S(T)$ at time $T \leq T_p < T_1$. It is hence more naturally valued in T_p-forward measure, even though the market-implied distribution of $S(T)$ is known in the annuity measure. We have

$$V_{CMS}(0) = A(0)E^A\left[\frac{P(T,T_p)}{A(T)}S(T)\right] \tag{108}$$

Lemma 8.1 (Replication method for CMS pay-offs)**.** *Assuming that the annuity mapping function $\alpha(S(T)) = \frac{P(T,T_p)}{A(T)}$ is known, then*

$$V_{CMS}(0) = A(0)S(0)\alpha(S(0)) + \int_{-\infty}^{S(0)} w(K)V_{rec}(0,K)\,dK + \int_{S(0)}^{\infty} w(K)V_{pay}(0,K)\,dK \tag{109}$$

where

$$\begin{aligned} V_{rec}(0,K) &= A(0)E^A\left[(K-S(T))^+\right] \\ V_{pay}(0,K) &= A(0)E^A\left[(S(T)-K)^+\right] \\ w(s) &= \frac{d^2}{ds^2}(\alpha(s)s) \end{aligned}$$

where $V_{rec}(0,K), V_{pay}(0,K)$ are the time-0 values of receiver and payer swaptions.

Using iterated conditioning one easily sees:

Proposition 8.2. *The annuity mapping function $\alpha(s)$ may be written as the conditional expectation*

$$\alpha(s) = E^A\left[\left.\frac{P(T,T_p)}{A(T)}\right| S(T) = s\right] \tag{110}$$

In order to compute the conditional expectation (110), it is useful to remember that given two random variables X, Y, the conditional expectation can be interpreted as

$$E[X|Y] = f^*(Y), \qquad f^* = \mathrm{argmin}\left\{E\left[(X-f(Y))^2\right], f \in \mathcal{B}\right\}$$

for a space $\mathcal{B}$ of suitably regular functions on Y, generally defined by a parametric functional form $\mathcal{B} = \left\{f(y;\theta), \theta \in \Theta \subset \mathbb{R}^d\right\}$. Imposing the conditions

$$\frac{\partial}{\partial \theta_i} E\left[(X - f(Y;\theta))^2\right] = 0, \quad i = 1, \dots, d$$

we obtain the following Proposition.

Proposition 8.3. *Given two random variables X and Y and a paremetric set of functions $\{f(y;\theta), \theta \in \Theta \subset \mathbb{R}^d\}$, an approximation to $E[X|Y]$ is given by*

$$E[X|Y] \simeq f(Y;\theta^*)$$

where θ^* is a solution to the set of equations

$$E\left[X\frac{\partial f}{\partial \theta_i}(Y;\theta)\right] = E\left[f(Y;\theta)\frac{\partial f}{\partial \theta_i}(Y;\theta)\right], \quad i = 1, \ldots, d \tag{111}$$

8.6.1 Linear TSR model

Assume $\alpha(s) = \alpha_1 s + \alpha_2$. Imposing the no-arbitrage condition we obtain

$$\frac{P(0,T_p)}{A(0)} = E^A\left[\frac{P(T,T_p)}{A(T)}\right] = E^A\left[\alpha_1 S(T) + \alpha_2\right] = \alpha_1 S(0) + \alpha_2$$

whence

$$\alpha_2 = \frac{P(0,T_p)}{A(0)} - \alpha_1 S(0)$$

α_1 can be considered an exogenous input. The value of the CMS payoff is hence

$$\begin{aligned} V_{CMS}(0) &= A(0)E^A\left[\frac{P(T,T_p)}{A(T)}S(T)\right] = E^A[\alpha(S(T))S(T)] = E^A\left[(\alpha_1 S(T) + \alpha_2)S(T)\right] \\ &= P(0,T_p)S(0) + \alpha_1 A(0)\mathrm{Var}^A[S(T)] \end{aligned}$$

where the variance $\mathrm{Var}^A[S(T)]$ can be computed (as in the Libor-in-arrears case) by the replication method[27]

$$\mathrm{Var}^A[S(T)] = 2\int_{-\infty}^{S(0)} S(K)p(0,S(0);T,K)\,dK + 2\int_{S(0)}^{\infty} S(K)c(0,S(0);T,K)\,dK$$

8.6.2 Libor market model

Establishing dependency explicitly would be too complicated if our sole purpose is to price CMS cash-flows, but it would be useful for applications of Libor market models to exotic derivatives that are linked to CMS rates (e.g. callable CMS range accruals or CMS spread TARN's).

Proposition 8.4. *See [AP10-3, Proposition 16.6.3]. The annuity mapping function (8.2) in the Libor market model is approximated by*

$$\alpha(s) = E^A\left[\frac{P(T,T_p)}{A(T)}\middle| S(T) = s\right] \simeq \left(\sum_{n=0}^{N-1} \tau_n \prod_{i=0}^{n} \frac{1}{1+\tau_i \ell_i(s)}\right)^{-1}$$

where

$$\ell_n(s) = L_n(0)\left(1 + c_n\frac{s - S(0)}{S(0)}\right)$$

[27] Recall that for any twice-differentiable function we have

$$\begin{aligned} E[f(S(T))] = f(K^*) + f'(K^*)(S(0) - K^*) + \int_{-\infty}^{K^*} p(0,S(0);T,K)f''(K)\,dK \\ + \int_{K^*}^{\infty} c(0,S(0);T,K)f''(K)\,dK \end{aligned} \tag{112}$$

In our case, we apply it with $K^* = S(0)$ and $f(S(T)) = (S(T) - S(0))^2$.

8.6.3 Quasi-Gaussian model

Recall the bond reconstitution formula in the qG-model

$$\begin{aligned} P(T,M) &= P(T,M;x(T),y(T)), \\ P(T,M;x,y) &= \frac{P(0,M)}{P(0,T)} \exp\left(-G(T,M)x - \frac{1}{2}G(T,M)^2 y\right) \end{aligned}$$

and recall that $y(T)$ is well approximated by a deterministic function

$$\bar{y}(T) = \left(\frac{\partial L}{\partial x}(T,0,0)\right)^{-2} \mathrm{Var}^A(S(T))$$

where $\mathrm{Var}^A(S(T))$ is computed consistently with the market via replication as above. One can then define an annuity mapping

$$\alpha(s)\frac{P(T,T_p;X(T,s),\bar{y}(T))}{A(T,X(T,s),\bar{y}(T))}$$

where $X(T,s)$ is the inverse function (in x) to $S(T,x,\bar{y}(T))$ which can then be used to compute $V_{CMS}(0)$ via (111).

Computational remark. Note that

$$\begin{aligned} \frac{d\alpha}{ds} = \frac{1}{A(T,X(T,s),\bar{y}(T))} &\left[\frac{\partial P(T,T_p,X(T,s),\bar{y}(T))}{\partial x}\frac{\partial X(T,s)}{\partial s}A(T,X(T,s),\bar{y}(T))\right. \\ &\left. -P(T,T_p,X(T,s),\bar{y}(T))\frac{\partial A(T,X(T,s),\bar{y}(T))}{\partial x}\frac{\partial X(T,s)}{\partial s}\right] \end{aligned}$$

To compute $\frac{d\alpha(K)}{ds}$ we need

- $X(T,K)$: by construction $S(T,X(T,s),\bar{y}(T)) = s$ so $X(T,K)$ is the x-solution to the equation $S(T,x,\bar{y}(T)) = K$.
- $\frac{\partial X(T,K)}{\partial s}$: differentiating implicitly we obtain

$$X(T,S(T,x,\bar{y}(T))) = x \Longrightarrow \frac{\partial X(T,s)}{\partial s}\frac{\partial S(T,x,\bar{y}(T))}{\partial x} = 1$$

whence

$$\frac{\partial X}{\partial s}(T,K) = \left(\frac{\partial S}{\partial x}(T,K,\bar{y}(T))\right)^{-1}$$

One proceeds analogously to compute $\frac{d^2\alpha(K)}{ds^2}$.

8.6.4 Correcting non-arbitrage-free methods

Some annuity mapping methods are not arbitrage-free by construction (e.g. Libor) and in some others, like the linear TSR model, despite being arbitrage-free theoretically, numerical methods can induce slight errors. We introduce a modification to amend this error. The valuation formula for CMS cash-flows reads

$$V_{CMS}(0) = A(0)E^A\left[\frac{P(T,T_p)}{A(T)}S(T)\right] = A(0)E^A\left[\alpha(S(T))S(T)\right]$$

In order to have an arbitrage-free model, since $\frac{P(T,T_p)}{A(T)}$ is a tradeable asset, we need

$$E^A\left[\alpha(S(T))\right] = E^A\left[\frac{P(T,T_p)}{A(T)}\right] = \frac{P(0,T_p)}{A(0)}$$

Since this may fail in some models, we can re-scale the function $\alpha(s)$ to force the previous equality to be satisfied. Defining

$$\tilde{\alpha}(s) = \frac{P(0,T_p)}{A(0)} \frac{\alpha(s)}{E^A\left[\alpha(S(T))\right]},$$

it is clear that $E^A\left[\tilde{\alpha}(S(T))\right] = \frac{P(0,T_p)}{A(0)}$ and we obtain the *improved* CMS valuation formula

$$V_{CMS}(0) = A(0)E^A\left[S(T)\tilde{\alpha}(S(T))\right] = P(0,T_p)\frac{E^A\left[S(T)\alpha(S(T))\right]}{E^A\left[\alpha(S(T))\right]}$$

This correction is useful even for arbitrage-free models, since there may be arbitrage induced by the numerical scheme.

8.6.5 CDF and PDF of CMS rate in Forward measure

We look at an alternative to the replication method of Lemma 8.1, since the weights may be hard to compute for discontinuous payoffs. The problem of pricing a cash-flow that pays $g(S(T))$ at time $T_p \geq T$ can be formulated as

$$E^{T_p}\left[g(S(T))\right] = \int_{-\infty}^{\infty} g(s)\psi^{T_p}(s)\,ds$$

where $\psi^{T_p}(s)$ is the density of $S(T)$ in the T_p-forward measure.

The CDF and PDF of the swap rate $S(T)$ in the annuity measure[28] are given respectively by

$$\Psi^A(K) = 1 + \frac{\partial}{\partial K}c(K), \qquad \psi(K) = \frac{\partial^2}{\partial K^2}c(K)$$

where $c(K) = E^A\left[(S(T)-K)^+\right]$. Changing to the T_p-forward measure we obtain

Proposition 8.5. *Given an annuity mapping function* $\alpha(s) = E^A\left[\left.\frac{P(T,T_p)}{A(T)}\right| S(T) = s\right]$, *the PDF* $\psi^{T_p}(s)$ *and the CDF* $\Psi^{T_p}(s)$ *of the swap rate* $S(T)$ *in the* T_p*-forward measure are given by*

$$\psi^{T_p}(s) = \frac{A(0)}{P(0,T_p)}\alpha(s)\psi^A(s), \tag{113}$$

$$\Psi^{T_p}(s) = \frac{A(0)}{P(0,T_p)}\int_{-\infty}^{s}\alpha(u)\psi^A(u)\,du \tag{114}$$

Sketch of proof. Note that $\psi^{T_p} = E^{T_p}\left[\delta(S(T)-K)\right]$, so changing to the annuity measure and using iterated conditioning yields the result. □

When the annuity mapping function is linear $\alpha(s) = \alpha_1 s + \alpha_2$, expressions (113-114) take a simple form

[28] The price of a call option is given by $C(S,K) = \int_K^\infty \varphi_{S(T)}(x)(x-K)\,dx$ Differentiating with respect to K twice we obtain

$$\begin{aligned}\frac{\partial C(S,K)}{\partial K} &= \frac{\partial}{\partial K}\left[\int_K^\infty \varphi_{S(T)}(x)\,dx - K\int_K^\infty \varphi_{S(T)}(x)\,dx\right] = -\int_K^\infty \varphi_{S(T)}(x)\,dx \\ \frac{\partial^2 C(S,K)}{\partial K^2} &= \varphi_{S(T)}(K)\end{aligned}$$

Corollary 8.6. *In the linear TSR model, namely when $\alpha(s) = \alpha_1 s + \alpha_2$, the CDF $\Psi^{T_p}(s)$ and the PDF $\psi^{T_p}(s)$ of the swap rate in the T_p-forward measure are given by*

$$\begin{aligned} \Psi^{T_p}(s) &= \frac{A(0)}{P(0,T_p)} \left[\alpha_1(S(0) - s - c(s)) + \alpha(s)\Psi^A(s)\right] & (115) \\ \psi^{T_p}(s) &= \frac{A(0)}{P(0,T_p)} (\alpha_1 s \alpha_2)\psi^A(s), & (116) \\ c(s) &= E^A\left[(S(T) - s)^+\right] \end{aligned}$$

Finally notice that since a CMS caplet payoff is given by

$$V_{CMS-caplet}(0,K) = P(0,T_p)E^{T_p}\left[(S(T) - K)^+\right],$$

there is an explicit link between $V_{CMS-caplet}(0,K)$ and $\psi^{T_p}(s)$.

Lemma 8.7 (Link between CMS caplet payoff and swap rate in T_p-forward measure)**.** *The market-implied PDF $\psi^{T_p}(s)$ of the swap rate in the T_p-forward measure can be obtained from values of CMS caplets by*

$$\psi^{T_p}(K) = \frac{1}{P(0,T_p)} \frac{\partial^2}{\partial K^2} V_{CMS-caplet}(0,K) \qquad (117)$$

8.6.6 SV model for CMS rate

It is useful, specially when dealing with vanilla derivatives linked to multiple market rates, to have PDFs and CDFs thereof in the forward measure to be of tractable form. Unfortunately, this is not the case if one starts with the SV model in the annuity measure and then switches to the forward measure. We seek to obtain an approximation to the PDF $\psi^{T_p}(s)$ that is from the same family as the PDF $\psi^A(s)$.

Assume the swap rate follows

$$\begin{aligned} dS(t) &= \lambda\left(bS(t) + (1-b)S(0)\right)\sqrt{z(t)}dW^A(t) & (118) \\ dz(t) &= \theta(1 - z(t))dt + \eta\sqrt{z(t)}dZ^A(t), \quad z(0) = 1 & (119) \\ E[dW^A(t)dZ^A(t)] &= 0 \end{aligned}$$

This system defines the PDF $\psi^A(\cdot)$ of $S(T)$ in measure Q^A. Define an adjusted process $\tilde{S}(t)$ given by the following dynamics in the T_p-forward measure

$$\begin{aligned} d\tilde{S}(t) &= \tilde{\lambda}\left(\tilde{b}\tilde{S}(t) + (1-\tilde{b})\tilde{S}(0)\right)\sqrt{\tilde{z}(t)}dW^{T_p}(t) \\ d\tilde{z}(t) &= \tilde{\theta}(1 - \tilde{z}(t))dt + \tilde{\eta}\sqrt{\tilde{z}(t)}dZ^{T_p}(t), \quad \tilde{z}(0) = 1 & (120) \\ E[dW^{T_p}(t)dZ^{T_p}(t)] &= 0 \end{aligned}$$

Align the mean of $\tilde{S}(T)$ to equal the CMS-adjusted value of $S(T)$, namely

$$\tilde{S}(0) = E^{T_p}[S(T)]$$

Goal: define $\psi^{T_p}(\cdot)$ to be the PDF of $\tilde{S}(T)$ given by the previous system of SDEs and aim to set the adjusted parameters $\tilde{\lambda}, \tilde{b}, \tilde{\eta}$ such that the distribution of $\tilde{S}(T)$ is as close as possible to that of $S(T)$ in the T_p-forward measure.

- $\tilde{S}(T) = E^{T_p}[S(T)]$ is calculated by the replication method of Lemma 8.1.

- Since $V_{CMS-caplet}(0,K) = P(0,T_p)E^{T_p}\left[(S(T)-K)^+\right]$ and $\tilde{S}(0) = E^{T_p}[S(T)]$, by iterated conditioning we see that in fact

$$V_{CMS-caplet}(0,K) = P(0,T_p)E^{T_p}\left[(\tilde{S}(T)-K)^+\right]$$

so that CMS caplets are nothing more than European calls on $\tilde{S}(T)$. Therefore, $\tilde{\lambda}, \tilde{b}, \tilde{\eta}$ can be obtained by direct calibration of the SV model (120) (which is in the T_p-forward measure) to prices of CMS caplets that are computed in the original SV model[29] for $S(T)$ (with dynamics specified in the annuity measure).

8.6.7 Dynamics of CMS rate in forward measure

Consider a stochastic volatility model in the Q^A-measure

$$\begin{aligned} dS(t) &= \lambda\varphi(S(t))\sqrt{z(t)}dW^A(t) && (121)\\ dz(t) &= \theta(1-z(t))dt + \eta\psi(z(t))dZ^A(t), \quad z(0)=1 && (122)\\ E[dW^A(t)dZ^A(t)] &= 0 \end{aligned}$$

In the previous subsection we absorbed the measure change to Q^{T_p} into the SV diffusion parameters: namely, we defined an analogous system in the $\tilde{Q}^{T_p}$-measure and adjusted its parameters to obtain market-consistent results. We now seek to bring the two-dimensional SDE (121)-122) into an approximation of the T_p-forward measure[30]. Concretely, we construct a measure $\tilde{Q}^{T_p}$ such that for any function $f(\cdot)$ we have

$$E^A\left[\frac{P(T,T_p)}{A(T)}f(S(T))\right] = \frac{P(0,T_p)}{A(0)}\tilde{E}^{T_p}[f(S(T))]$$

Proposition 8.8. *See [AP10-3, Proposition 16.6.8]. Defining the measure $\tilde{Q}^{T_p}$ by the condition that the process $(z(t),S(t))$ satisfies the following SDE system in $\tilde{Q}^{T_p}$*

$$\begin{aligned} dS(t) &= \lambda\varphi(S(t))\sqrt{z(t)}\left(d\tilde{W}^{T_p}(t) + v^S(t)dt\right) && (123)\\ dz(t) &= \theta(1-z(t))dt + \eta\psi(z(t))\left(d\tilde{Z}^{T_p}(t) + v^Z(t)\right), \quad z(0)=1 \\ E\left[dW^{T_p}dZ^{T_p}\right] &= 0 \end{aligned}$$

where the drift adjustments are given by

$$\begin{aligned} v^Z(t) &= \eta\psi(z(t))\partial_z \ln\left[\Lambda(t,z(t),S(t))\right] && (124)\\ v^S(t) &= \lambda\varphi(S(t))\sqrt{z(t)}\partial_S \ln\left[\Lambda(t,z(t),S(t))\right] \end{aligned}$$

and where the function $\Lambda(t,z,s)$ satisfies the following PDE

$$\partial_t\Lambda(t,z,s) + \theta(1-z)\partial_z\Lambda(t,z,s) + \frac{1}{2}\eta^2\psi(z)^2\partial^2_{zz}\Lambda(t,z,s) + \frac{1}{2}\lambda^2\varphi(s)^2 z\partial^2_{ss}\Lambda(t,z,s) = 0, \quad t\in[0,T] \quad (125)$$

with terminal condition

$$\Lambda(T,z,s) = \hat{\alpha}(s) := \frac{A(0)}{P(0,T_p)}\alpha(s) \quad (126)$$

[29] These are obtained by the usual replication method of Lemma 8.1.

[30] Namely, a measure that will resemble a forward measure for European-type payoffs fixing at T and paying at T_p.

Then for any function $f(\cdot)$ we have

$$\frac{A(0)}{P(0,T_p)}E^A\left[\frac{P(T,T_p)}{A(T)}f(S(T))\right] = \tilde{E}^{T_p}[f(S(T))]$$

Remark 8.1. This Proposition establishes a numerical scheme for simulating $(z(t), S(t))$ in $\tilde{Q}^{T_p}$ for the purpose of pricing European-style derivatives fixing at T and paying at T_p.

1. Solve the PDE (125-126) numerically on a grid (t, z, s).
2. Perform Monte Carlo simulation for (123), using the drift adjustments $v^Z(t)$ and $v^S(t)$, which are computed for each path using (124).

Remark 8.2. If $\hat{\alpha} = \hat{\alpha}_1 s + \hat{\alpha}_2$ then no finite difference scheme is needed since

$$\begin{aligned} dv^z(t) &= 0, \\ dv^S(t) &= \lambda\varphi(S(t))\sqrt{z(t)}\frac{\hat{\alpha}_1}{\hat{\alpha}_1 S(t) + \hat{\alpha}_2} \end{aligned}$$

8.7 Quanto CMS

Given a swap rate $S(T)$ in domestic currency, a quanto CMS cash-flow pays $g(S(T))$ at time $T_p \geq T$ in some other foreign currency and it thus has value

$$V_{QuantoCMS}(0) = \beta_f(0)E^f\left[\frac{1}{\beta_f(T_p)}g(S(T))\right]$$

Let $X(t)$ be the foreign exchange rate expressing one unit of domestic currency in foreign currency units. In the domestic risk-neutral measure (and in foreign currency units) the pricing formula can be written as

$$\begin{aligned} V_{QuantoCMS}(0) &= \frac{\beta_d(0)}{X(0)}E^d\left[\frac{1}{\beta_d(T_p)}g(S(T))X(T_p)\right] \\ &\overset{[1]}{=} \frac{\beta_d(0)}{X(0)}E^d\left[\frac{1}{\beta_d(T)}g(S(T))\beta_d(T)E^d_T\left[\frac{1}{\beta_d(T_p)}X(T_p)\right]\right] \\ &\overset{[2]}{=} \frac{\beta_d(0)}{X(0)}E^d\left[\frac{1}{\beta_d(T)}g(S(T))P(T,T_p)X_{T_p}(T)\right] \end{aligned}$$

where in [1] we used iterated conditioning and in [2] we used a change of measure $\beta_d(T)E^d_T\left[\frac{1}{\beta_d(T_p)}X(T_p)\right] = P_d(T,T_p)E^d_T[X(T_p)]$ and the notation $X_{T_p}(T) = \frac{P_f(T,T_p)X(T)}{P_d(T,T_p)}$.

Changing to the annuity measure we obtain

$$V_{QuantoCMS}(0) = \frac{A(0)}{X(0)}E^d\left[\frac{1}{A(T)}g(S(T))P_d(T,T_p)X_{T_p}(T)\right] = \frac{A(0)}{X(0)}E^{A,d}\left[g(S(T))v(S(T))\right]$$

where

$$v(s) = E^{A,d}\left[\frac{P_d(T,T_p)}{A(T)}X_{T_p}(T)|S(T)=s\right].$$

Recalling the definition of the annuity mapping function $\alpha(s) = E^{A,d}\left[\frac{P(T,T_p)}{A(T)}|S(T)=s\right]$ and defining $\chi(s) = E^{A,d}\left[X_{T_p}(T)|S(T)=s\right]$ we can approximate

$$v(s) \simeq \alpha(s)\chi(s)$$

so that
$$V_{QuantoCMS}(0) \simeq \frac{A(0)}{X(0)} E^{A,d}\left[g(S(T))\alpha(S(T))\chi(S(T))\right]$$
and this can be computed via replication once the function $\chi(\cdot)$ is determined.

In order to compute $\chi(s) = E^{A,d}\left[X_{T_p}(T)|S(T) = s\right]$ one needs to specify a joint distribution of the swap rate $S(T)$ and the forward FX rate $X_{T_p}(T)$ and this can be accomplished by means of a Gaussian copula. The marginal one-dimensional PDF of $S(T)$ in measure $Q^{A,d}$ is given by $\psi^A(s)$ (given by the swaption model). Assuming that $X_{T_p}(T)$ is log-normal in $Q^{A,d}$ we have
$$X_{T_p}(T) = X(0)e^{m_X T + \sigma_X \sqrt{T}\xi_1}, \quad \xi_1 \sim N(0,1).$$
where

- σ_X is obtained by calibrating to T-expiry ATM options on the FX rate. For long-dates quanto contracts, the FX smile make start to matter and we may have to allow σ_X to vary stochastically. This can be accomplished, for instance, using copulas.
- m_X is clarified below.

Given $\xi_2 \sim N(0,1)$ with $\rho_{XS} = \mathrm{Corr}(\xi_1, \xi_2)$ we have
$$S(T) \overset{d}{=} (\Psi^A)^{-1}\left(\Phi(\xi_2)\right)$$
whence
$$\begin{aligned}
\chi(s) &= E^{A,d}\left[X_{T_p}(T)|S(T) = s\right] \\
&= X(0)e^{m_X T} E^{A,d}\left[e^{\sigma_X\sqrt{T}\xi_1}|\xi_2 = \Phi^{-1}(\Psi^A(s))\right] \\
&\overset{not}{=} X(0)e^{m_X T}\tilde{\chi}(s), \\
\tilde{\chi}(s) &= \exp\left(\rho_{XS}\sigma_X\sqrt{T}\mathcal{N}^{-1}(\Psi^A(s)) + \frac{1}{2}\sigma_X^2 T(1-\rho_{XS}^2)\right)
\end{aligned}$$

In order to establish the constant m_X we impose a no-arbitrage condition. Since $X_{T_p}(\cdot)$ is a $Q^{T_p,d}$-martingale we have
$$\begin{aligned}
X_{T_p}(0) &= E^{T_p,d}[X_{T_p}(T)] = \frac{A(0)}{P_d(0,T_p)} E^{A,d}\left[\frac{P_d(T,T_p)}{A(T)} X_{T_p}(T)\right] \\
&= \frac{A(0)}{P_d(0,T_p)} E^{A,d}\left[\alpha(S(T))\chi(S(T))\right] \\
&= X(0)e^{m_X T}\frac{A(0)}{P_d(0,T_p)} E^{A,d}\left[\alpha(S(T))\tilde{\chi}(S(T))\right]
\end{aligned}$$
and we can then solve for $e^{-m_X T}$.

8.8 Eurodollar futures

Eurodollar futures are exchange-traded contracts on Libor rates. The latter are inputs into an interest rate curve construction whereas the former are liquidly quoted in the market. It is important to be able to value ED futures efficiently, due to their high trading volume and their use in yield curve construction, which rules out Monte Carlo methods and thus necessitates the development of analytic approximations that incorporate the volatility smile.

Daily mark-to-market causes forward and futures contracts values to differ. If $F(t;T,M)$ and $L(t;T,M)$ are the time-t futures and Libor rates respectively covering the period $[T,M]$, recall that $L(t;T,M) = E_t^M[L(T;T,M)]$.

For futures contracts that are marked-to-market continuously we have

$$F(t;T,M) = E_t[L(T;T,M)]$$

If mark-to-market happens discretely, consider a tenor structure $0 = T_0 < T_1 < \cdots < T_N$ and let $L_n(t) = L(t;T_n,T_{n+1})$. Denote by Q^B the spot Libor measure induced by the process

$$B(t) = P(t,T_{q(t)}) \prod_{n=0}^{q(t)-1} (1 + \tau_n L_n(T_n))$$

Then if the futures rate is marked-to-market only on the dates $T_0 < T_1 < \cdots < T_n = T < M$, we have

$$F(t,T;M) = E_t^B[L(T;T,M)]$$

8.8.1 The problem and the plan

Assuming that all futures rates $\{F_n(0)\}_{n=0,\ldots,N-1}$ are known (market-observable), our goal is to derive formulas that express forward Libor rates $\{L_n(0) = L(0;T_n,T_{n+1})\}$ in terms of futures $\{F_n(0)\}$. The road map to derive the ED futures valuation formula is:

1. Use expansion technique to express a forward rate as a (model-independent) function of a collection of futures rates with expiries on or before the expiry of the forward rate.
2. Separate variance terms appearing in previous formula into slow- and fast-moving.
3. Fast-moving volatility parameters
4. Slow-moving correlation parameters

8.8.2 Expansion around futures value

Since $L_n(0) = E^{n+1}[L_n(T_n)]$ for all $n = 0,\ldots,N-1$, changing to the spot measure we obtain

$$L_n(0) = E^B\left[\frac{dQ^{n+1}}{dQ^B} L_n(T_n)\right] = E^B\left[L_n(T_n)\prod_{i=0}^{n}\frac{1+\tau_i L_i(0)}{1+\tau_i L_i(T_i)}\right] \tag{127}$$

We seek to express the expectation in the RHS of (127) in terms of market-observed quantities by expanding it in powers of a small parameter that measures the deviation of each $L_n(T_n)$ from its mean in the spot measure $F_n(0) = E^B[L_n(T_n)]$. Define

$$\begin{aligned} L_n^\epsilon(t) &\stackrel{def}{=} F_n(0) + \epsilon\,(L_n(t) - F_n(0)) \\ V(\epsilon) &= L_n^\epsilon(T_n)\prod_{i=0}^{n}\frac{1+\tau_i L_i(0)}{1+\tau_i L_i^\epsilon(T_i)} \end{aligned}$$

Since clearly $L_n^1(t) = L_n(t)$, we have $L_n(0) = E^B[V(1)]$. Expanding $V(\epsilon)$ into a Taylor series in ϵ yields

$$V(\epsilon) = V(0) + E^B\left[\frac{dV}{d\epsilon}(0)\right]\epsilon + \frac{1}{2}E^B\left[\frac{d^2V}{d\epsilon^2}(0)\right]\epsilon^2 + \mathcal{O}(\epsilon^3) \tag{128}$$

Computing these derivatives (c.f. [AP10-3, Lemma 16.8.4]) one obtains the following:

Theorem 8.9. *For any $n = 0, \ldots, N-1$, an approximation to the forward rate $L_n(0)$ is obtained from the futures $\{F_i(0)\}_{i=0}^{n}$ and forward rates for previous periods $\{L_i(0)\}_{i=0}^{n-1}$ by sequentially solving the equations*

$$L_n(0) = E^B[V(1)] = V(0)\left(1 + \frac{1}{2}\sum_{j,m=0}^{n} D_{j,m}\, Cov^B(L_j(T_j), L_m(T_m))\right) \tag{129}$$

with

$$\begin{aligned} V(0) &= F_n(0)\prod_{i=0}^{n}\frac{1+\tau_i L_i(0)}{1+\tau_i F_i(0)} \\ D_{j,m} &= \cdots \end{aligned}$$

8.8.3 Forward rate variances

The covariances in (129) are, by definition

$$\mathrm{Cov}^B(L_j(T_j), L_m(T_m)) = \sqrt{\mathrm{Var}^B(L_j(T_j))}\sqrt{\mathrm{Var}^B(L_m(T_m))}\mathrm{Corr}^B(L_j(T_j), L_m(T_m))$$

The spot-measure variances $\mathrm{Var}^B(L_n(T_n))$ can be approximated in the forward measure[31]

$$\mathrm{Var}^B(L_n(T_n)) \simeq \mathrm{Var}^{n+1}(L_n(T_n))$$

The variances $\mathrm{Var}^{n+1}(L_n(T_n))$ can be computed by calibrating a low-parametric vanilla model to liquid strikes. Assuming

$$\begin{aligned} dL_n(t) &= \sigma_n\left(b_n L_n(t) + (1-b_n)L_n(0)\right)\sqrt{z(t)}dW^{n+1}(t) \\ dz(t) &= \theta\left(1 - z(t)\right)dt + \eta_n\sqrt{z(t)}dZ^{n+1}(t) \\ E\left[dW^{n+1}(t)dZ^{n+1}(t)\right] &= 0 \end{aligned}$$

After calibrating the parameters (σ_n, b_n, η_n) to swaption prices, the forward variances are given by

$$\begin{aligned} \mathrm{Var}^{n+1}(L_n(T_n)) &= \frac{L_n^2(0)}{b_n^2}\left[\Psi_{\bar{z}}((\sigma_n b_n)^2, 0; T_n)\right] \\ \Psi_{\bar{z}}(v, u; t) &= \cdots \end{aligned}$$

8.8.4 Forward rate correlations

See [AP10-3, Page 759].

[31] See [AP10-3, Page 757]. The error made by the approximation is claimed to be small.

9 Multi-rate vanilla derivatives

Here we look at payoffs that are linked to more than one swap rate, the CMS spread being the most prominent example, without resorting to a full term structure model. Concretely, given a payment date T_p, a collection of swap or Libor rates $S_1(t), \ldots, S_d(t)$, a collection of fixing dates $t_1, \ldots, t_d$ and a d-argument function $f(s_1, \ldots, s_d)$, a multi-rate derivative is defined by a time-T_p payoff

$$V(T_p) = f(S_1(t_1), \ldots, S_d(t_d))$$

We outline the notion of copula and we indicate how it can be used to specify and control the dependence structure between the rates involved.

9.1 Dependence structure via copulas

The *copula approach* is a method of constructing a joint distribution of random variables consistently with pre-specified one-dimensional marginal distributions.

9.2 Gaussian copula method

Assume we are given one-dimensional CDF's $\Psi_1(\cdot), \ldots, \Psi_d(\cdot)$ and that we want to construct a multi-dimensional random vector $(X_1, \ldots, X_d)$ with a measure of control over the dependence of the random variables X_i, constrained so that each variable X_i has CDF $\Psi_i(\cdot)$. We can accomplish this by specifying $X_i = \Psi_i^{-1}(\Phi(Z_i)$ where $(Z_1, \ldots, Z_d)$ is a multi-dimensional normalized Gaussian vector with correlation matrix R and where $\Phi(\cdot)$ is the standard one-dimensional Gaussian CDF. Clearly

$$P(X_i \le x) = P(\Psi_i^{-1}(\Phi(Z_i) \le x) = P(Z_i \le \Phi^{-1}(\Psi_i(x))) = \Phi\left(\Phi^{-1}(\Psi_i(x))\right) = \Psi_i(x)$$

so X_i indeed has CDF $\Psi_i(\cdot)$. Denote by

- $\Psi_{gauss}(x_1, \ldots, x_d)$ the CDF of $(X_1, \ldots, X_d)$ just constructed.
- $\Phi_d(z_1, \ldots, z_d; R)$ the CDF of $(Z_1, \ldots, Z_d)$.

Then

$$\begin{aligned}
\Psi_{gauss}(x_1, \ldots, x_d) &= P\left(X_1 \le x_1, \ldots, X_d \le x_d\right) \\
&= P\left(Z_1 \le \Phi^{-1}(\Psi_1(x)), \ldots, Z_d \le \Phi^{-1}(\Psi_d(x))\right) \\
&= \Phi_d\left(\Phi^{-1}(\Psi_1(x_1)), \ldots, \Phi^{-1}(\Psi_d(x_d)); R\right) \qquad (130)
\end{aligned}$$

One easily obtains the PDF as

$$\begin{aligned}
\psi(x_1, \ldots, x_d) &= \frac{\partial^d}{\partial x_1 \cdots \partial x_d} \Psi_{gauss}(x_1, \ldots, x_d) \\
&= \phi_d\left(\Phi^{-1}(\Psi_1(x_1)), \ldots, \Phi^{-1}(\Psi_d(x_d)); R\right) \cdot \prod_{i=1}^{d} \frac{\psi_i(x_i)}{\phi\left(\Phi^{-1}(\Psi_i(x_i))\right)}
\end{aligned}$$

where $\phi(z)$ and $\phi_d(z_1, \ldots, z_d; R)$ are the one- and d-dimensional Gaussian PDF's, and $\psi_i(x_i)$ is the one-dimensional PDF of X_i.

9.3 General couplas

Note that we can re-write the Gaussian copula CDF (130) as

$$\begin{aligned}\Psi_{gauss}(x_1,\ldots,x_d) &= \Phi_d\left(\Phi^{-1}(\Psi_1(x_1)),\ldots,\Phi^{-1}(\Psi_d(x_d));R\right)\\ &= C_{gauss}\left(\Psi_1(x_1),\ldots,\Psi_d(x_d);R\right)\end{aligned}$$

where

$$C_{gauss}(u_1,\ldots,u_d;R)=\Phi_d\left(\Phi^{-1}(u_1),\Phi^{-1}(u_d);R\right)$$

is a multi-dimensional distribution function for a vector of d uniform random variables[32].

Definition 9.1 (general copula)**.** *A d-dimensional copula function is a function $C:[0,1]^d\to[0,1]$, $C(u_1,\ldots,u_d)$ defining a valid joint distribution function for a d-dimenaional vector of random variables, with each variable being uniformly distributed on $[0,1]$.*

Requiring that $C(u_1,\ldots,u_d)$ be a distribution imposes strong constraints on the form of the function C. In particular

$$\begin{aligned}C(u_1,u_2,\ldots,u_{i-1},0,u_{i+1},\ldots,u_d) &= 0\\ C(1,1,\ldots,1,u_i,1,\ldots,1) &= u_i\end{aligned}$$

The simplest examples of couplas are the following:

1. *Independence copula*: if all d uniform random variables underlying the copula are independent, we obtain
$$C_{ID}(u_1,\ldots,u_d)=\prod_{i=1}^{d}u_i$$
2. *Perfect dependence copula*: introduce d uniform random variables $U_1=\cdots=U_d$. Then
$$C_D(u_1,\ldots,u_d)=P(U_1\le u_1,\ldots,U_d\le u_d)=P\left(U_1\le\min_{i=1,\ldots,d}u_i\right)=\min_{i=1,\ldots,d}u_i$$

The following result allows us to construct copulas from other copulas and will be the basis to obtain a useful copula for multi-rate derivatives.

Proposition 9.2 (making copulas from copulas)**.** *Let $C_1,\ldots,C_M$ be d-dimensional copulas and let $q_{m,i}:[0,1]\to[0,1]$, $m=1,\ldots,M$, $i=1,\ldots,d$ be functions that are either strictly increasing or identically equal to 1. Suppose that $\prod_{m=1}^{M}q_{m,i}(u)=u$ for $u\in[0,1]$ and $\lim_{u\to0^+}q_{m,i}(u)=q_{m,i}(0)$. Then*

$$C(u_1,\ldots,u_d)=\prod_{m=1}^{M}C_m(q_{m,1}(u_1),\ldots,q_{m,d}(u_d))$$

is a coupla

In particular, taking $M=2$, $C_1=\prod_{i=1}^{d}u_i$ the independence copula, $C_2=C_{gauss}(u_1,\ldots,u_d;R)$, $q_{1,i}(u)=u^{1-\theta_i}$, $q_{2,i}(u)=u^{\theta_i}$ for $\theta_i\in[0,1]$, $i=1,\ldots,d$ we obtain the so-called *power Gaussian copula.*

Corollary 9.3 (power Gaussian copula)**.** *Let R be a $d\times d$ correlation matrix and $\theta=(\theta_1,\ldots,\theta_d)\in[0,1]^d$ a d-dimensional vector of parameters. Then the power Gaussian function*

$$C_{PG}(u_1,\ldots,u_d;R)=\left(\prod_{i=1}^{d}u^{1-\theta_1}\right)C_{gauss}(u_1^{\theta_1},\ldots,u_d^{\theta_d};R) \tag{131}$$

[32] Recall that if X is a random variable with CDF $F_X(x)$ then $Y=F_X(X)$ has a uniform distribution over $(0,1)$. Indeed,

$$P\left[Y\le y\right]=P\left[F_X(X)\le y\right]=P\left[X\le F_X^{-1}(y)\right]=F_X\left(F_X^{-1}(y)\right)=y$$

9.4 Copula methods for spread options

Spread options are the most liquid among the multi-rate vanilla derivatives. The payoff of a spread option is given by $(S_1(T) - S_2(T) - K)^+$ at time T_p, so the (undiscounted) value f the spreads options is

$$V(0; T, K) = E^{T_p}\left[(S_1(T) - S_2(T) - K)^+\right]$$

Using a two-dimensional Gaussian copula with correlation

$$\rho = \mathrm{Corr}(S_1(T), S_2(T), \qquad R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

1. Derive market-implied density for each swap rate $S_i(T)$ under its own annuity measure $\psi_i^{A_i}(s)$, $i = 1, 2$ using swaption prices[33].
2. Use Proposition 8.5 to convert each $\psi_i^{A_i}(x)$ into the PDF $\psi_i^{T_p}(x)$ in the T_p-forward measure

$$\psi^{T_p}(s) = \frac{A(0)}{P(0,T_p)}\alpha(s)\psi^A(s), \tag{132}$$

$$\Psi^{T_p}(s) = \frac{A(0)}{P(0,T_p)}\int_{-\infty}^{s}\alpha(u)\psi^A(u)\,du \tag{133}$$

where $\alpha(s) = E^A\left[\left.\frac{P(T,T_p)}{A(T)}\right| S(T) = s\right]$.

3. Given the correlation matrix R, use the Gaussian copula function to construct a joint PDF

$$\psi^{T_p}(x_1, x_2; R) = c_{gauss}(\Psi_1^{T_p}(x_1), \Psi_2^{T_p}(x_2)) \cdot \psi_1^{T_p}(x_1)\psi_2^{T_p}(x_2)$$

where

$$c_{gauss}(u_1, u_2; R) = \frac{\phi_2\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2); R\right)}{\phi(\Phi^{-1}(u_1))\phi(\Phi^{-1}(u_2))}$$

4. Integrate the payoff $(x_1 - x_2 - K)^+$ against the density $\psi^{T_p}(x_1, x_2; R)$ to obtain the undiscounted model price of the spread option

$$V_{mdl}(0; T, K, R) = \int\int (x_1 - x_2 - K)^+\psi^{T_p}(x_1, x_2; R)\,dx_1\,dx_2$$

Finally, in order to calibrate the model to market:

5. Find the implied spread correlation $\rho = \rho(T, K)$ such that

$$V_{mdl}(0; T, K, R) = V_{mkt}(0; T, K)$$

The main issue with this algorithm stems from the so-called *correlation smile*: for a fixed maturity date, one obtains a different implied spread correlation for each strike. A simple Gaussian copula has insufficient flexibility to capture the market-observed distributions of CMS spreads. Essentially, varying ρ allows one to shift the implied spread option volatility smile in parallel, making it impossible to match the market-implied smile (see figure 2).

[33] Explain

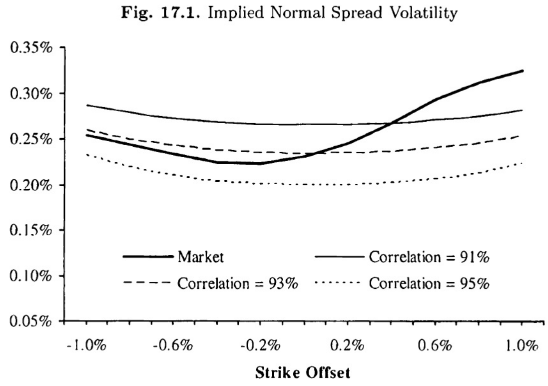

Figure 2: Implied normal spread volatility for a 5-year spread option on the difference between 10-year and 2-year swap rates, [AP10-3, Figure 17.1].

9.4.1 Power gaussian

One way of fixing the correlation smile issue outlined in the previous subsection is to resort to a two-dimensional power Gaussian copula

$$C_{PG}(u, v; \rho, \theta_1, \theta_2) = u^{1-\theta_1} v^{1-\theta_2} C_{gauss}(u^{\theta_1}, v^{\theta_2}; \rho)$$

where ρ is a correlation coefficient and $\theta_1, \theta_2 \in [0, 1]$. As above, observes empirically that ρ allows to shift the implied volatility spread vertically, while the parameters θ_1, θ_2 provide good control over the slope and curvature of the smile (see figure 3).

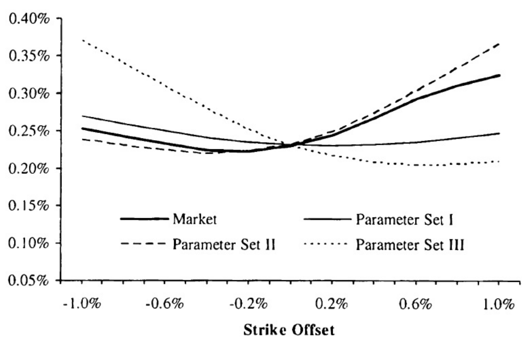

Figure 3: Implied normal spread volatility for a 5-year spread option on the difference between 10-year and 2-year swap rates obtained using the power Gaussian copula in three parameter scenarios, [AP10-3, Figure 17.2].

9.5 Limitations of the copula method

- Copulas do not result n a dynamic model for the yield curve.
- Copulas allow us to easily ascribe a joint terminal distribution to a collection of CMS rates, for the purpose of pricing European multi-rate options.
- Copulas in practical use are not usually chosen for their links to observed financial relationships but instead for their technical properties and ease of implementation.

- Even though the power Gaussian copula is capable of reproducing a wide range of market-observed shapes of spread volatility smiles, to date no copula has been found that reproduced market volatilities of spread options exactly for all values of spread strikes. See [AP10-3, Section17.4.4] for some insights.

10 Callable Libor Exotics

10.1 Review of basic notions

- In an *exotic swap*, a regular floating Libor rate is swapped against structured coupons that are allowed to be arbitrary functions of observed interest rates (e.g. Libor or CMS rates).

  For instance, *floor can be represented as an exotic swap* in which a Libor rate is exchanged for a floored payoff.

$$(k - L_n(T_n))^+ = (k - L_n(T_n))^+ + L_n(T_n) - L_n(T_n) = \max(k, L_n(T_n)) - L_n(T_n)$$

- [AP10-1, Section 5.13] Exotic swaps start their life as bonds or notes, sold by banks to investors. The investor pays an upfront principal amount to the note issuer, who in turn pays the investor a structured coupon, and repays the principal at the maturity of the note. The principal amount is invested by the issuer and pays de Libor rate, plus or minus a spread. From the perspective f the issuer, the net cash-flows of the note are those of an exotic swap.
- If C_n is the structured coupon for the n-th period, the value of the exotic swap equals (from perspective of structured leg buyer)

$$V_{exotic}(t) = \beta(t) \sum_{n=0}^{N-1} \tau_n E_t \left[\frac{1}{\beta(T_{n+1})} (C_n - L_n(T_n)) \right]$$

- C_n can be a complicated function of interest rates, structured to reflect investors' views on the market.
- A *Bermudan swaption* is an option to enter into a fixed-floating swap on any of its fixing dates (or a subset of them). Once exercised on date T_n, say, the holder enters the swap with the first fixing date T_n and final payment date T_N

  At time T_n, the value of a payer swap, if exercised, is

$$U_n(T_n) = \beta(T_n) \sum_{i=n}^{N-1} \tau_i E_{T_n} \left[\frac{1}{\beta(T_{i+1})} (L_i(T_i) - k) \right]$$

  A Bermudan contract swaption is hence an option to choose between $U_n(T_n)$ for $n = 0, \ldots, N-1$ and its time T_n value is

$$V_{bermudan-swaption}(T_n) = \max\{U_n(T_n), H_n(T_n)\}$$

  where $H_n(T_n))$ is the value of a Bermudan swaption with exercise dates $\{T_i\}_{i=n+1}^{N-1}$.
- A *callable Libor exotic* (CLE) is a Bermudan-style option to enter an exotic swap. The difference with respect to an ordinary exotic swap is that the issuer of the note has the right to cancel (or *call*) the note on a schedule of dates. The principal is then returned to the investor and no further coupons are paid.
- Given an exotic swap with structured coupons $\{C_n\}_{n=0}^{N-1}$. Its time-T_n value if exercised is

$$U_n(T_n) = \beta(T_n) \sum_{i=n}^{N-1} \tau_i E_{T_n} \left[\frac{1}{\beta(T_{i+1})} (C_i - L_i(T_i)) \right]$$

  The n-th hold value of the CLE, denoted $H_n(T_n))$ is the time-T_n value of the CLE provided it has not been exercised on or before T_n.

10.2 Model calibration for CLE's

CLE's present non-trivial dependencies on the dynamics of market rates and generally require sophisticated term-structure models.

The values of exotic derivatives are not directly observable in the market and therefore models for these securities have to be calibrated indirectly to other market information that is deemed relevant for the class of exotics under consideration.

There are three sources of potentially relevant information:

1. *Market prices of liquid vanilla interest rate derivatives (or market-implied - spot - volatilities) of various strikes.* Complex structured products can be decomposed into simpler liquid derivatives. For instance, a *callable inverse floater* with exercise dates $T_1 < T_2 < \cdots < T_{N-1}$ is a Bermudan option to swap a Libor rate and a structured coupon

$$C_n = \min\{\max\{6\% - L_n(T_n), 0\%\}, 4\%\}$$

and it can be decomposed as

$$C_n = (6 - L_n(T_n))^+ - (2 - L_n(T_n))^+$$

namely a portfolio of a long floorlet struck at 6% and a short floorlet struck at 2%. The corresponding market-implied volatilities should thus be included as targets for model calibration.

2. *Volatilities of core / coterminal rates.* Even though the underlying exotic swap has no dependencies on spot volatilities other than those underlying the floorlets inherent in the CIF, the callability structure introduces additional dependence on other vanilla derivatives.

   For instance, the callable inverse floater will depend on the implied volatility of the swap rate that fixes on T_i and runs for the period $[T_i, T_N]$ for each $i = 1, \ldots, N-1$.

3. *Historical information about market quantities*, such as volatilities and correlations of market rates. In the callable inverse floater, assuming that the option hasn't been exercised by time T_n, the time-T_n value of the remaining part of the underlying swap depends on caplet volatilities observed at time T_n (which are only known at time T_n and are called *forward volatilities.*

   Any model will impose certain dynamics on the volatility structure of interest rates, and one should make sure that the model projections for forward volatilities are in line with market-implied information.

4. Modeler's belief of what constitutes reasonable behavior of the model parameters.

10.3 Valuation theory

We use the spot Libor measure Q^B, with numraire $B(t)$ defined on a tenor structure $0 = T_0 < T_1 < \cdots < T_N$. The net payment seen by the structured coupon receiver at time T_{n+1} is $X_n = \tau_n(C_n - L_n(Tn))$.

- The n-th *exercise value*, namely the value of all future payments if the CLE is exercised at time T_n is

$$U_n(t) = B(t) \sum_{i=n}^{N-1} E_t\left[\frac{X_i}{B(T_{i+1})}\right], \qquad U_N(t) \equiv 0$$

- The n-th *hold value* is the value of a CLE where the exercise opportunities have been restricted to $\{T_{n+1}, \ldots, T_{N-1}\}$

$$H_n(T_n) = B(T_n) \sup_{\xi \in \mathcal{T}_n} E_{T_n}\left[\frac{U_\xi(T_\xi)}{B(T_\xi)}\right]$$

which can be computed recursively

$$\begin{aligned} H_{n-1}(T_{n-1}) &= B(T_{n-1}) E_{T_{n-1}}\left[\frac{1}{B(T_n)} \max\{H_n(T_n), U_n(T_n)\}\right], \qquad (134) \\ H_{N-1} &\equiv 0 \end{aligned}$$

If

$$\eta_n(\omega) = \min\{k > n : U_k(T_k) > H_k(T_k)\} \wedge N$$

then

$$H_0(0) = E\left[\frac{U_{\eta_0}(T_{\eta_0})}{B(T_{\eta_0})}\right] = E\left[\sum_{n=\eta_0}^{N-1} \frac{X_n}{B(T_{n+1})}\right]$$

10.4 Monte Carlo valuation

The regression-based least squares scheme relies on the assumption that the hold values $H_n(T_n)$ can be regressed against simulated state variables of the model at time T_n.

We fix some notation first. Consider:

- $\zeta(t) = (\zeta_1(t), \ldots, \zeta_q(t))^\top$ q-dimensional vector process of regression variables.
- X random variable, with value $X(\omega)$ on a Monte Carlo path ω.
- $\omega_1, \ldots, \omega_K$ K simulated Monte Carlo paths.
- $\mathcal{R}_T(X)$ results of regression of the K-dimensional vector $(X(\omega_1), \ldots, X(\omega_K))$ on the $K \times q$ matrix $\zeta(T)^\top$ of regression variable observations at time T, namely

$$\mathcal{R}_T(X) = \zeta(T)^\top \beta, \qquad [\zeta(T)]_{i,j} = \zeta_i(T, \omega_j)$$

where

$$\beta = \arg\min_\beta \left\| \left(X(\omega_1) - \zeta(T,\omega_1)^\top \beta, \ldots, X(\omega_K) - \zeta(T,\omega_K)^\top \beta\right) \right\|^2$$

The most basic least squares scheme for CLE valuation consists of replacing the conditional expected value operator $E_{T_{n-1}}$ in (134) with the regression operator $\mathcal{R}_{T_{n-1}}$, namely

$$\begin{aligned} \tilde{H}_{n-1}(T_{n-1}) &= \mathcal{R}_{T_{n-1}} \overbrace{\left[\frac{B(T_{n-1})}{B(T_n)} \max\left\{\tilde{H}_n(T_n), U_n(T_n)\right\}\right]}^{V_n(\omega_k)}, \quad n = N-1, \ldots, 1 \qquad (135) \\ \tilde{H}_{N-1} &\equiv 0 \end{aligned}$$

The full algorithm[34] is the following:

1. Choose and calibrate a term structure model (e.g. LM).
2. Choose regression variables $\zeta(t)$.
3. Simulate K paths $\omega_1, \ldots, \omega_K$, each one representing (in the LM case) one simulated path of all core Libor rates.

[34] See [AP10-3, Page 824].

4. For each path ω_k, calculate $B(T_n, \omega_k)$, $n = 1, \ldots, N-1$.
5. For each path, calculate the exercise value $U_n(T_n, \omega_k)$ of the underlying exotic swap on all exercise dates $n = 1, \ldots, N-1$ (more work may be needed here if these are not available in closed form, see Section 10.4.1 below).
6. For each path ω_k, calculate the values of the q-dimensional regression variables $\zeta(T_n, \omega_k)$, $n = 1, \ldots, N-1$.
7. Set $\tilde{H}_{N-1} \equiv 0$.
8. For each $n = N-1, \ldots, 1$
   (a) Form a K-dimensional vector $V_n = (V_n(\omega_1), \ldots, V_n(\omega_K))^\top$,
   $$V_n(\omega_k) = \frac{B(T_{n-1}, \omega_k)}{B(T_n, \omega_k)} \max \left\{ \tilde{H}_n(T_n, \omega_k), U_n(T_n, \omega_k) \right\}, \quad k = 1, \ldots, K$$
   (b) Calculate
   $$\tilde{H}_{n-1}(T_{n-1}) = \mathcal{R}_{T_{n-1}}(V_n) = \zeta(T_{n-1})^\top \beta$$
   by regressing the vector V_n against the matrix of regression variables observed on date T_{n-1}, namely by finding a q-dimensional
   $$\beta = \arg\min_\beta \left\| \left(V_n(\omega_1) - \zeta(T_{n-1}, \omega_1)^\top \beta, \ldots, V_n(\omega_K) - \zeta(T_{n-1}, \omega_K)^\top \beta \right) \right\|^2$$
9. $\tilde{H}_0(T_0)$ is our sought for estimate of the value of the CLE.

The scheme presents several shortcomings:

(i) The exercise values $U_n(T_n)$ may not me computable in closed form, and they are necessary to evaluate (135) at each step.

(ii) The use of nested regression (we apply $\mathcal{R}_{T_n}$ to a function of a regressed value $\tilde{H}_n(T_n)$) could spawn significant biases.

(iii) It is unclear whether the $\tilde{H}_n$'s are low- or high-biased estimates of H_n's.

10.4.1 Regression for the underlying

The above scheme can be easily extended to arbitrary underlying swaps, which is useful if the latter are too complex and do not admit a closed form. The idea is to apply regression to compute $U_n(T_n)$ using

$$U_n(t) = \sum_{i=n}^{N-1} E_t \left[\frac{B(t)}{B(T_{i+1})} X_i \right], \quad X_n = \tau_n(C_n - L_n(T_n))$$

1. For each path ω_k and each fixing date T_n calculate the net coupons $X_n(T_n, \omega_k)$.
2. Use regression to estimate the exercise value $U_n(T_n)$ as
   $$\tilde{U}_n(T_n) = \mathcal{R}_{T_n} \left(\sum_{i=n}^{N-1} \frac{B(T_n)}{B(T_{i+1})} X_i \right) = \zeta(T_n)^\top \beta$$
   where
   $$\beta = \arg\min_\beta \left\| \left(\sum_{i=n}^{N-1} \frac{B(T_n)}{B(T_{i+1})} X_i(\omega_1) - \zeta(T_n, \omega_1)^\top \beta, \ldots, \sum_{i=n}^{N-1} \frac{B(T_n)}{B(T_{i+1})} X_i(\omega_K) - \zeta(T_n, \omega_K)^\top \beta \right) \right\|^2$$

This can be computed recursively as

$$\begin{aligned}\tilde{U}_n(T_n) &= \mathcal{R}_{T_n}(Y_n), \\ Y_n &= \frac{B(T_n)}{B(T_{n+1})}(X_n + Y_{n+1}), \quad n = N-1, \ldots, 1\end{aligned}$$

3. Replace the expression $V_n(\omega_k)$ inside of $\mathcal{R}_{T_{n-1}}$ in (135) by

$$V_n(\omega_k) = \frac{B(T_{n-1})}{B(T_n)} \max\left\{\tilde{H}_n(T_n), \tilde{U}_n(T_n)\right\}$$

10.4.2 Using regressed variables for decision only

The values that are regressed at time T_{n-1} themselves come as the result of a regression at time T_n. Compounded regression can lead to substantial biases, even though these can be reduced significantly by using regressed variables for decision-making purposes only, as explained in [AP10-3, Section 18.3.4] and as we summarize next.

Define $G_n(t) = H_n(t) - U_n(t)$ and note (c.f. [AP10-3, Equation 18.18]) that

$$G_{n-1}(T_{n-1}) = E_{T_{n-1}}\left[\frac{B(T_{n-1})}{B(T_n)}\left(-X_{n-1} + G_n(T_n)^+\right)\right], \quad n = N, \ldots, 0$$

Hence $G_0(0)$ is the value of a swap paying coupons $-X_n$ on dates T_n, plus the right to cancel it on any of the exercise dates (CLE).

This provides an *alternative way of valuing the CLE*: we can compute a least squares approximation of $G_0(0)$ and obtain the value of the CLE by subtracting the swap value from it. The least squares version of the latter (i.e. replacing $E_{T_{n-1}}$ by $\mathcal{R}_{T_{n-1}}$ thus reads

$$\begin{aligned}\tilde{G}_{n-1}(T_{n-1}) &= \mathcal{R}_{T_{n-1}}(V_n), \qquad \tilde{G}_{N-1}(T_{N-1}) \equiv 0, \\ V_n(\omega_k) &= \frac{B(T_{n-1}, \omega_k)}{B(T_n, \omega_k)}\left(-X_{n-1}(\omega_k) + \tilde{G}_n(T_n, \omega_k)^+\right), \quad k = 1, \ldots, K\end{aligned}$$

We next show how to bypass the need to use compounded regression: note that

$$\begin{aligned}G_0(0) &= H_0(0) - U_0(0) = -E\left[\sum_{n=0}^{\eta-1} \frac{X_n}{B(T_{n+1})}\right] = -E\left[\sum_{n=0}^{N-1} \frac{X_n}{B(T_{n+1})} \mathbb{I}_{\{\eta > n\}}\right] \\ &= -E\left[\sum_{n=0}^{N-1} \frac{X_n}{B(T_{n+1})}\left(\prod_{i=1}^{n} \mathbb{I}_{\{G_i(T_i) > 0\}}\right)\right]\end{aligned}$$

Defining

$$V_n = \frac{B(T_n)}{B(T_{n+1})}\left[-X_n + \mathbb{I}_{\{G_{n+1}(T_{n+1}) > 0\}} V_{n+1}\right], \quad n = N-1, \ldots, 0$$

we obtain

$$V_0 = -\sum_{n=0}^{N-1} \frac{X_n}{B(T_{n+1})}\left(\prod_{i=1}^{n} \mathbb{I}_{\{G_i(T_i) > 0\}}\right)$$

whence taking expectations

$$G_0(0) = E[V_0]$$

Critically, the recursion for V_n *involves the value of the cancelable note for exercise decisions only*, whereas the coupon values are never regressed. We can thus manage to avoid compounded regression by defining the following approximation:

$$\begin{aligned}\hat{G}_{n-1}(T_{n-1}) &= \frac{B(T_{n-1})}{B(T_n)}\left[-X_{n-1} + \mathbb{I}_{\{\tilde{G}_n(T_n)>0\}}\hat{G}_n(T_n)\right],\\ \tilde{G}_{n-1}(T_{n-1}) &= \mathcal{R}_{T_{n-1}}(\hat{G}_{n-1}(T_{n-1}))\end{aligned}$$

10.4.3 Controlling the MC estimate bias

So far the bias of our MC estimate for the value of a CLE is unknown:

- The exercise decisions used in the schemes are necessarily suboptimal, which suggests that the estimate is biased low.
- However, the previous schemes use the same set of sample paths to estimate the exercise decision as to calculate the security value if exercised, which can lead to an upward bias (since future information can affect our decision to exercise).

In order to remedy this and guarantee a low bias, one could use the LS regression to estimate the exercise decision rule only, and then use an independent simulation to calculate the value of the CLE given that exercise rule. Concretely:

1. Run any of the regression schemes above to produce coefficients (betas) for the hold $\mathcal{C}(\tilde{H}_n(T_n))$ and exercise values $\mathcal{C}(\tilde{U}_n(T_n))$ at all exercise times, $n = 1, \ldots, N-1$.
2. Simulate K' additional paths $\omega'_1, \ldots, \omega'_{K'}$ that are independent from the paths used in the regression scheme.
3. For each path ω'_k, calculate the values of the q-dimensional regression variables process $\zeta(T_n, \omega'_k)$, $n = 1, \ldots, N-1$.
4. For each path ω'_k calculate an estimate of the exercise index $\tilde{\eta}$ by

$$\tilde{\eta}(\omega'_k) = \min\left\{n \geq 1 : \mathcal{C}(\tilde{U}_n(T_n))^\top \zeta(T_n, \omega'_k) \geq \mathcal{C}(\tilde{H}_n(T_n))^\top \zeta(T_n, \omega'_k)\right\} \wedge N$$

5. Calculate the CLE value as the MC value of a knock-in discrete barrier option

$$H_0(0) \sim \frac{1}{K'}\sum_{k=1}^{K'}\left(\sum_{n=\tilde{\eta}(\omega'_k)}^{N-1} \frac{1}{B(T_{n+1}, \omega'_k)} X_n(\omega'_k)\right) \tag{136}$$

10.4.4 Upper bound

10.4.5 Choice of regression variables

The closer $\zeta(T) = (\zeta_1(T), \ldots, \zeta_q(T))$ approximates the information in $\mathcal{F}_T$ that is relevant for the security, the better the regression method performs (namely, the smaller the bias of the lower bound estimates of the security value).

- The dimensionality of the state vector (i.e. Libor rates) is typically so large that the regression will suffer from numerical problems. This issue can be handled by forming the first d principal components of the vector of Libor rates. Concretely, given a vector a (centered) still-alive forward Libor rates

$$A = (L_n(T_n) - E[L_n(T_n)], \ldots, L_{N-1}(T_n) - E[L_{N-1}(T_n)])^\top, \quad N - n \geq d$$

one can estimate the term covariance matrix of the rates $c = E\left[AA^\top\right]$ and by PCA we can find an $(N-n)\times d$ matrix D such that $DD^\top$ is the closest rank-d approximation to c, namely

$$D = \arg\min_D \left[\mathrm{tr}(c - DD^\top)(c - DD^\top)^\top\right]$$

and then

$$A \simeq Dx, \qquad \text{where } x = \arg\min_x \|A - Dx\|^2 = (D^\top D)^{-1}D^\top A$$

- In some situations, the information carried in the state variables is inadequate for the security in question and in other situations the state variables may carry too much information, adding unnecessary work to the numerical scheme and reducing the quality of the CLE price estimate. See [AP10-3, Section 18.3.9.2].

10.5 Implementation

[AP10-3, Section 18.3.10] explores a number of ways in which the regression algorithm can be made more robust. Recall that given a K-dimensional vector $Y = (Y_1, \ldots, Y_K)^\top$ of simulated values of a random variable Y and a $K \times q$ matrix Z of simulated values of the regression variables $\zeta(t)$, namely $Z_{k,j} = \zeta_j(t, \omega_k)$, our goal in regression is to find a q-dimensional vector β such that solves the minimization problem

$$\beta = \arg\min_\beta \|Y - Z\beta\|^2$$

This problem has a naive solution $\beta = (Z^\top Z)^{-1}Z^\top Y$.

- [AP10-3, Section 18.3.10.1] explains the R^2 criterion for automated explanatory variable selection.

- [AP10-3, Section 18.3.10.4] explains two methods to make the computation $\beta = (Z^\top Z)^{-1}Z^\top Y$ more robust when the matrix $Z^\top Z$ is ill-conditioned, namely close to singular.

  (i) Tikhonov regularization: consider the minimization problem $\beta = \arg\min_\beta \|Y - Z\beta\|^2 + \omega_{reg}\|\beta\|^2$ where the scalar regularization weight ω_{reg} can be selected in a number of ways, for instance

  $$\omega_{reg} = \epsilon\sqrt{\frac{1}{q}\mathrm{tr}(Z^\top Z Z^\top Z)}, \qquad \epsilon = 10^{-4}$$

  The solution is given by

  $$\beta = (Z^\top Z + \omega_{ref} I)^{-1}Z^\top Y$$

  where the matrix to be inverted $Z^\top Z + \omega_{ref} I$ always has full rank.

  (ii) Singular value decomposition. Write $\beta = (Z^\top Z)^{-1}Z^\top Y$ as a system of linear equations

  $$M\beta = Z^\top Y, \qquad M = Z^\top Z$$

  and decompose the $q \times q$ matrix $M = U\Sigma V^\top$ where...

10.6 Valuation with low-dimensional models

Single-rate CLE's[35] can be valued accurately using the so-called *local-projection method*, which consists of calibrating a *local* low-dimensional model to the volatility information that has been identified as important to the CLE valuation.

As a starting point, any low-dimensional model should be calibrated to the following targets:

1. Underlying swap rate volatilities $\{S_n^1(T_n)\}_{n=1}^{N-1}$. Here we assume that each coupon C_n depends on $S_n^1(T_n)$ only and that $S_n(t) \equiv S_{n,\mu(n)}(t)$ where $\mu(n)$ is the number of periods.
2. Core swap rate volatilities $\{S_n^2(T_n)\}_{n=1}^{N-1}$, namely $S_n^2(t) \equiv S_{n,N-n}(t)$. We have flexibility in choosing strikes, but ATM strikes are common.
3. Core correlations for $\{S_n^2(T_n)\}_{n=1}^{N-1}$. These are *not observable in the market*, so we need to draw on an LM model which has been calibrated to the market as a whole. This is paramount: by including this dynamic information, the local model not only captures the static information about the interest rate volatilities at valuation time, but also the dynamics of the volatility structure.

There are several options for the local model, the one-dimensional quasi-Gaussian (qG) model being a natural one. Recall that a standard Gaussian model is given by

$$\begin{aligned}
dr(t) &= \kappa(t)\left[\theta(t) - r(t)\right] dt + \sigma_r(t) dW(t) \\
df(t,T) &= \sigma_f(t,T)^t \left[\int_t^T \sigma_f(t,u)\, du\right] dt + \sigma_f(t,T)^t dW(t) \\
\sigma_f(t,T) &= \sigma_r(t) \exp\left(-\int_t^T \kappa(u)\, du\right) \\
\sigma_f(t,T) &= g(t)h(T)
\end{aligned}$$

where one normally works with the state variables

$$\begin{aligned}
x(t) &:= r(t) - f(0,t) \\
dx(t) &= \left[y(t) - \kappa(t)x(t)\right] dt + \sigma_r(t) dW(t), \quad x(0) = 0, \\
y(t) &= \int_0^t e^{-2\int_u^t \kappa(s)\, ds} \sigma_r^2(u)\, du
\end{aligned}$$

In a *local volatility quasi-Gaussian (qG) model* the function $g(\cdot)$ is allowed to depend on the state variables, namely $g(t) \equiv g(t, x(t), y(t))$. The approximate dynamics of a local volatility quasi-Gaussian model are given by

$$\begin{aligned}
dS(t) &= \varphi(t, S(t)) dW^A(t) \\
\varphi^2(t,s) &= E^A\left[\{\partial_x S(t,x(t),y(t))\sigma_r(t,x(t),y(t))\}^2 \,|S(t) = s\right]
\end{aligned}$$

where

$$\varphi(t, S(t)) \approx \lambda_S(t)\left[b_S(t)S(t) + (1 - b_S(t))S(0)\right]$$

and

$$\lambda_r(t) = \sum_{n=1}^{N-1} \lambda_n 1_{(T_{n-1},Tn]}(t), \quad b_r(t) = \sum_{n=1}^{N-1} b_n 1_{(T_{n-1},Tn]}(t)$$

In this model we have;

[35] Namely, CLE's for which each coupon C_n depends on at most a single market rate, which we denote by $S_n^1(T_n)$.

- $2(N-1)$ independent volatility parameters (λ_n, b_n), $n = 1, \ldots, N-1$ which can be used to calibrate the model to term volatilities for one of the swap rate strips.
- $N-1$ parameters κ_n, $n = 1, \ldots, N-1$ defining the mean reversion which can be used to either match the term volatilities for the second swap rate strip or the core correlations.

Therefore, the one-factor local volatility qG model is not large enough to match all three sets of calibration targets identified above. Even though this is acceptable for some securities, it is not for some others, and in this case we might need to resort to a two-factor model.

10.7 Suitable analogue for core swap rates

We noted that the callability value is driven by the volatilities of core swap rates $S_n^2(t) = S_{n,N-n}(t)$, since CLE's are related to standard Bermudan swaptions. In some cases though, it is more relevant to calibrate the model to other European swaptions[36]. We thus try to refine the selection of the volatility targets relevant for the callability option of a CLE.

We use the idea is that the local model should match the values of *European options on exercise values* $U_n(T_n)$, $n = 1, \ldots, N-1$. Since such options can be hard to come by, we can linearize the underlying $U_n(T_n)$ of the CLE and *use the resulting rate as a replacement for the core swap rate.*

We use an LM model as a backdrop. Assume that

$$\begin{aligned} U_n(T_n) &= f_n(\mathbf{L}(T_n)), \quad n = 1, \ldots, N-1 \\ \mathbf{L}(T_n) &= (L_n(T_n), \ldots, L_{N-1}(T_n)) \end{aligned}$$

Linearizing we obtain

$$U_n(T_n) = f_n(\mathbf{L}(0)) + \nabla f_n(\mathbf{L}(0))[\mathbf{L}(T_n) - \mathbf{L}(0)]$$

so the value of the European option on the underlying can be approximated by

$$E\left[\frac{(U_n(T_n))^+}{B(T_n)}\right] \approx E\left[\frac{1}{B(T_n)}\left(\nabla f_n(\mathbf{L}(0))\mathbf{L}(T_n) - (\nabla f_n(\mathbf{L}(0))\mathbf{L}(0) - f_n(\mathbf{L}(0)))\right)^+\right]$$

and the *relevant interest rate* is hence

$$R_n(T_n) := \nabla f_n(\mathbf{L}(0))\mathbf{L}(T_n) = \sum_j \omega_{n,j} L_j(T_n), \quad \omega_{n,j} = \frac{\partial f_n}{\partial L_j}(\mathbf{L}(0))$$

- The volatility of the rate $R_n(T_n)$ can be approximated in an LM or local market models, so they can easily be used as volatility targets in place of core swap rates.
- $U_n(t)$ typically consists of options on market rates, so the derivatives $\frac{\partial f_n}{\partial L_j}$ can be computed with Black-type approximations of option values.

Example 10.1 (Bermudan swaptions)**.** The exercise value of a Bermudan swaption on a payer swap is

$$U_n(T_n) = \sum_{i=n}^{N-1} \tau_i (L_i(T_n) - K) \overbrace{\left(\prod_{k=n}^{i} \frac{1}{1 + \tau_k L_k(T_n)}\right)}^{P(T_n, T_{i+1})}$$

[36] For instance, in the case of Bermudan swaptions on amortizing swaps, the most relevant European swaptions are not the standard core European swaptions, but instead swaptions with tenors based on the durations of the underlying amortizing swap (see section ?? ahead).

with

$$
\begin{aligned}
f_n(\mathbf{x}) &= \sum_{i=n}^{N-1} \tau_i (x_i - K) \left(\prod_{k=n}^{i} \frac{1}{1+\tau_k x_k} \right) \\
\frac{\partial f_n}{\partial L_j}(\mathbf{x}) &= \tau_j \prod_{k=n}^{j} \frac{1}{1+\tau_k x_k} - \frac{\tau_j}{1+\tau_j x_j} \sum_{i=j}^{N-1} \tau_i (x_i - K) \left(\prod_{k=n}^{i} \frac{1}{1+\tau_k L_k(T_n)} \right)
\end{aligned}
$$

so

$$
\omega_{n,j} = \tau_j P(0; T_n, T_{j+1}) \left[1 - \frac{U_j(0)}{P(0, T_j)} \right]
$$

and $R_n(T_n) = \sum_{j=n}^{N-1} \omega_{n,j} L_j(T_n)$ is quite close to the core swap rate $S_{n,N-n}(T_n)$.

11 Bermudan swaptions

Given a tenor structure $0 = T_0 < T_1 < \cdots < T_{N-1}$, a Bermudan swaption is CLE with the coupon paying a fixed rate $C_n = k$ for $n = 1, \ldots, N-1$. Assume for simplicity that exercise is possible on any tenor date. If exercised at time T_n, the value for a payer swap in is

$$\begin{aligned} U_n(t) &= \sum_{i=n}^{N-1} P(t, T_{i+1}) E_t^{T_{i+1}} \left[\tau_i (L_i(T_i) - k)\right] = \sum_{i=n}^{N-1} \tau_i P(t, T_{i+1}) [L_i(t) - k]; \\ U_n(T_n) &= A_n(T_n) [S_n(T_n) - k] \end{aligned}$$

where $A_n(t) \equiv A_{n,N-n}(t)$ and $S_n(t) \equiv S_{n,N-n}(t)$.

Bermudan swaptions are liquid and their trading volume is high, so the performance advantage of PDE methods over Monte Carlo simulation make low-factor Markovian models attractive, and the *local projection method for single-rate CLE's* provides a sound framework for this. The method takes a simple form for Bermudan swaptions since only the volatility parameters of the core swap rates are relevant.

11.1 Local projection method

A Bermudan swaption gives the holder the right to exercise into one of several different swaps that are observed on different exercise dates. Bermudan swaptions are thus conceptually similar to a basket option with payoff $\max\{L_n(T_n)\}$, $n = 1, \ldots, N-1$ and their values are hence driven by the volatilities of core swap rates[37] $\{S_n(T_n)\}_{n=1}^{N-1}$ and by the inter-temporal correlations thereof[38].

This observation allows one to use simple one-factor Gaussian or quasi-Gaussian models for valuation ad risk-management of Bermudan swaptions, with the volatility function

[37] Recall that For each swap rate $S_{n,m}(t)$ we specify stochastic volatility (SV) dynamics in the annuity measure

$$\begin{aligned} dS_{n,m}(t) &= \lambda_{n,m} \left[b_{n,m} S_{n,m}(t) + (1 - b_{n,m}) S_{n,m}(0)\right] \sqrt{z_{n,m}(t)} dW^{n,m}(t) \\ dz_{n,m}(t) &= \theta \left[1 - z_{n,m}(t)\right] dt + \eta_{n,m} \sqrt{z_{n,m}} dZ^{n,m}(t) \end{aligned}$$

[38] Recall that in a one-factor pure Gaussian model or a quasi-Gaussian model (with local volatility), the mean reversion parameter can be calibrated to a market-implied inter-temporal correlation matrix

$$\hat{\chi}_{n_1,n_2} = \mathrm{Corr}(S_{n_1}(T_{n_1}), S_{n_2}(T_{n_2})), \quad 1 \le n_1 \le n_2 < N-1$$

independently of the volatility term. For instance in a pure Gaussian model with constant volatility σ_r and mean reversion κ we have (c.f. [AP10-2, Section 13.1.8.1])

$$\mathrm{Corr}(F(T_1; T_1, M_1), F(T_2; T_2, M_2)) = \mathrm{Corr}(x(T_1), x(T_2)) = e^{-\kappa(T_2 - T_1)} \left(\frac{1 - e^{-2\kappa T_1}}{1 - e^{-2\kappa T_2}}\right)^{1/2}$$

where $F(t; T, M) = -\frac{1}{M-T} \ln \frac{P(t,M)}{P(t,T)}$.

Likewise, in a one-factor quasi-Gaussian model we have

$$\begin{aligned} \mathrm{Corr}(S_{n_1}(T_{n_1}), S_{n_2}(T_{n_2})) \simeq & \int_0^{T_{n_1}} \left(\frac{\partial S_{n_1}}{\partial x}(t,0,0)\right) \left(\frac{\partial S_{n_2}}{\partial x}(t,0,0)\right) dt \\ & \times \left(\int_0^{T_{n_1}} \left(\frac{\partial S_{n_1}}{\partial x}(t,0,0)\right)^2 dt\right)^{-1/2} \times \left(\int_0^{T_{n_2}} \left(\frac{\partial S_{n_2}}{\partial x}(t,0,0)\right)^2 dt\right)^{-1/2} \end{aligned}$$

calibrated to volatilities of core rates and the mean-reversion function calibrated to intertemporal correlations thereof, with the latter being extracted from a global Libor model.

Several approaches are possible when it comes to choosing the calibration targets. Suppose we want to value and hedge a Bermudan swaption with fixed rate k and tenor structure $0 = T_0 < T_1 < \cdots < T_N$ giving the holder the right to exercise into one of the swaps observed on each of the exercise dates $T_1, \ldots, T_{N-1}$.

- For each expiry T_n, use the volatility of the appropriate ATM core European swaption. Unfortunately, this choice leads to inconsistent valuation between European and Bermudan swaptions: given a Bermudan swaption model calibrated to ATM European swaptions, applying it to a Bermudan swaption with a non-ATM strike and just a single exercise date (namely, a non-ATM European swaption) the value will differ from that obtained from that of the same derivative, priced as a European swaption.
- Instead, for each expiry T_n, use the volatility of the appropriate core European swaption corresponding to the strike k.

11.2 Amortizing, accreting and other non-standard Bermudan swaptions

Standard Bermudan swaptions can be extended by replacing the unit notional by a time-dependent and deterministic notional R_i, so that the exercise value becomes

$$\tilde{U}_n(t) = \sum_{i=n}^{N-1} R_i \tau_i P(t, T_{i+1})[L_i(t) - k] \tag{137}$$

- If the notional increases with the coupon index i, the Bermudan swaption is said to be *accreting.*
- If the notional decreases with the coupon index i, the Bermudan swaption is said to be *amortizing.*

Valuation of such swaptions in a properly calibrated model doesn't entail anything new. Nevertheless, while general models like LM present a product-independent calibration, models requiring local calibration, such as the one-factor quasi-Gaussian model, calibration for non-standard Bermudan swaptions requires a deeper analysis, since their liquidity is considerably poorer than for vanilla European swaptions.

We thus require a *pre-processing step to extract amortizing European swaption prices from a model calibrated to liquid vanilla European swaptions.*

11.2.1 Relating non-standard to standard swap rates

Consider a swap that starts at T_n and ends at T_N, with a notional schedule $\{R_i\}$ The annuity and swap rate that correspond to this non-standard swap are defined as

$$\tilde{A}_n(t) = \sum_{i=n}^{N-1} R_i \tau_i P(t, T_{i+1}), \qquad \tilde{S}_n(t) = \frac{1}{\tilde{A}_n(t)} \sum_{i=n}^{N-1} R_i \tau_i P(t, T_{i+1}) L_i(t)$$

The non-standard swap rate can be decomposed as a linear combination of standard swaps like so[39]:

[39] The weights $\omega_{n,m}(T_n)$ can be approximated reasonably well b their values at time 0.

$$\tilde{S}_n(t) = \sum_{m=1}^{N-1} \omega_{n,m}(T_n) S_{n,m}(T_n), \quad \omega_{n,m}(T_n) = (R_{n+m-1} - R_{n+m}) \frac{A_{n,m}(T_n)}{\tilde{A}_n(T_n)}$$

Hence in order to price a non-standard Bermudan swaption in principle one needs to calibrate the model to the volatilities of standard rates with all expiries $T_1, \ldots, T_{N-1}$ and all maturities, something which a low-dimensional local model will virtually never be able to do. We next discuss a method to remedy this issue.

11.2.2 Representative swaption approach (payoff matching)

The idea is to *choose a standard swap that approximates the non-standard swap* in some reasonable sense, and then to calibrate the Bermudan swaption model to the market-implied volatilities of swaptions on these standard swaps, one per exercise date.

Consider a one-factor Gaussian model to illustrate the idea. Fix a start date T_n and let $\tilde{U}_n(T_n; x)$ be the value of a non-standard swap, where $x = x(T_n)$ is the Gaussian short rate state on date T_n. We have seen that

$$\begin{aligned}
\tilde{U}_n(T_n; x) &= \sum_{m=1}^{N-m} (R_{n+m-1} - R_{n+m}) V_{n,m}(T_n) \\
&= \sum_{m=1}^{N-m} (R_{n+m-1} - R_{n+m}) \left[\sum_{i=n}^{n+m-1} \tau_i P(T_n, T_{i+1}) [L_i(T_n) - k] \right]
\end{aligned}$$

Recall that by the Gaussian bond reconstitution formula we have

$$\begin{aligned}
P(t,T) &= \frac{P(0,T)}{P(0,t)} \exp\left(-x(t) G(t,T) - \frac{1}{2} y(t) G^2(t,T) \right) \\
G(t,T) &= \int_t^T e^{-2\int_u^t \kappa(s)\,ds}\, du \\
y(t) &= \int_0^t e^{-2\int_u^t \kappa(s)\,ds} \sigma_r^2(u)\, du
\end{aligned}$$

which explains why we indicate the dependence on x in the notation $\tilde{U}_n(T_n; x)$. Note also that $\tilde{U}_n(T_n; x)$ depends on the volatility up to time T_n only.

We now seek to find a standard swaption providing the best possible approximation to $\tilde{U}_n(T_n; x)$. Let $V(x; R, q, m)$ denote the value of a standard swap starting on T_n as a function of x, with constant notional R, fixed rate (strike) q and covering m periods (where we allow m to be a real number in order to avoid discontinuities.

In the so-called *payoff matching* method, the three parameters R, q, m are chosen so that the level, slope and curvature of the swap payoffs as functions of the state variable x, namely we impose

$$\begin{aligned}
V(x_0; R, q, m) &= \tilde{U}_n(T_n; x_0) \\
\frac{\partial}{\partial x} V(x_0; R, q, m) &= \frac{\partial}{\partial x} \tilde{U}_n(T_n; x_0) \\
\frac{\partial^2}{\partial x^2} V(x_0; R, q, m) &= \frac{\partial^2}{\partial x^2} \tilde{U}_n(T_n; x_0)
\end{aligned} \tag{138}$$

where $x_0 = E[x(T_n)]$. These systems are solved numerically and produce the optimal parameters R_n^*, q_n^*, m_n^* (which are numerically seen to depend only mildly on mean reversion).

Since $\tilde{U}_n(T_n; x)$ only depends on the volatility parameter[40] $\sigma_r(t)$ over $[0, T_n]$, it is easy to incorporate the payoff matching methods into a volatility calibration scheme such as the one outlined in section 5.3.2. Specifically, we calibrate one swaption at a time:

1. Assume that the mean reversion has been fixed exogenously and that $\sigma_0, \ldots, \sigma_{i-1}$ have been computed.
2. Select the standard swaption $V_i(x; R_i^*, q_i^*, m_i^*)$ that best approximates $\tilde{U}_i(T_i; x)$ by solving the system of equations (138), which only requires knowledge of $\sigma_0, \ldots, \sigma_{i-1}$.
3. Set the value σ_i such that the model price for the standard swaption $V_i(x; R_i^*, q_i^*, m_i^*)$ equals its market price, by numerically inverting Jamshidian's formula (32)
4. Repeat Step 2 for $i = 0, \ldots, N-2$.

Remark 11.1. This method as is doesn't work well for accreting Bermudan swaptions, since it would produce an optimal tenor m^* that is longer than the tenor of the amortizing swap $N - n$ (see discussion on [AP10-3, Page 884]). It turns out that in order to get a reasonable calibration scheme for an accreting Bermudan swaption we would *need to calibrate to two standard European swaptions per expiry* and this can be difficult within a one-factor Markovian model, so additional factors are required.

11.2.3 Basket approach

This approach is a creative way of using one-factor models for non-standard Bermudan swaptions by splitting the valuation into two stages:

(i) Some model is used to calculate values of core non-standard European swaptions.

(ii) A one-factor model is then calibrated to the previous values and subsequently used to compute the value of non-standard Bermudan swaptions.

The variants of this method rely on different ways of implementing (i), namely of valuing non-standard European swaptions. One could for instance use a globally calibrated model such as LM to value non-standard European swaptions and then use the local projection method for non-standard Bermudan swaptions (namely, calibrate a low-dimensional model to the prices produced by the global model).

Alternatively, one may proceed as follows. Consider a swaption grid on a tenor $T_1 < \cdots < T_N$.

1. For each $n = 1, \ldots, N-1$, fit an instance of a one-factor Gaussian or quasi-Gaussian model to the prices of European swaptions expiring at T_n with underlying swaps spanning the intervals $[T_n, T_{n+1}], \ldots, [T_n, T_{N-1}]$.

   Note that the value of a swap fixing at T_n and covering m periods depends on the mean reversion function over $[T_n, T_{n+m}]$ so that we can keep σ_r fixed at a reasonable level (e.g. 1%) and calibrate the function $\kappa(t)$ one swaption at a time.

   Choice of European swaption strikes: there is no fixed rule for this. The best option is to use a one-factor quasi-Gaussian model with stochastic volatility, which allows calibration to the volatility smile at more than a single strike, thereby alleviating the strike selection problem.
2. For each T_n, use the relevant instance of the one-factor model calibrated in the previous step to price the T_n-expiry non-vanilla European option that the Bermudan swaption can be exercised into.

[40] In [AP10-3, Section 19.4.3] the authors claim that $\tilde{U}_n(T_n; x)$ does NOT depend on the volatility parameter, which is not true for the Gaussian model as far as I'm concerned. The one-swaption-at-a-time volatility calibration scheme can still be easily coupled with the payoff matching method though, since the time-T_n bond prices depend only on the volatility parameter up to time T_n.

3. Calibrate a one-factor model to the prices of the non-standard swaptions obtained in the previous step using the standard procedure (namely, calibrating the short rate volatility function $\sigma_r(t)$ one swaption at a time while keeping the mean reversion fixed at a user-specified value or at a level that makes inter-temporal correlations of core swap rates match those coming from a global model.

11.2.4 American swaptions

American swaptions allow the holder to exercise on any given date after the lockout period. These are particularly popular in the US for hedges of mortgage bonds.

American swaptions present a subtlety: if exercise takes place during $[T_n, T_{n+1}]$, the option holder receives a swap starting at T_{n+1} as well as an *exercise fee* equal to $(L_n(T_n) - k)(T_{n+1} - t)$, namely the exercise value per unit of notional is

$$U_n^A(t) = (L_n(T_n) - k)(T_{n+1} - t) + U_{n+1}(t)$$

Note that there time-t exercise value hence depends on the value of the Libor rate at $T_n < t$ and is thus *path-dependent.*

11.2.4.1 American vs. high-frequency Bermudan swaptions

11.2.4.2 The Proxy Libor rate method

11.3 Flexi-swaps

A *Bermudan swaption* can be interpreted as as fixed-floating swap with zero notional and a single option to increase the notional to a given level on any of the exercise dates. Similarly, a *cancelable swap* can be seen as a swap of full notional with an option to decrease the notional to zero on any of the exercise dates.

Flexi-swaps (a.k.a. *chooser swaps*) are contracts offering more flexibility in choosing swap notionals, namely swaps with multiple options to change the notional on a given set of exercise dates, subject to certain constraints.

Consider a tenor structure $\{T_n\}$, $n = 0, \ldots, N$ and a collection of coupons X_n with unit notional fixing at T_n and paying at T_{n+1}. A *flexi-swap* is a contract paying a net coupon $X_n R_n$ at time T_{n+1}, where R_0 is fixed up-front, and the time-T_n notional R_n is chosen by the holder of the option at time T_n, subject to some constraints:

- Global deterministic bonds: $R_n \in [g_n^{lo}, g_n^{hi}]$.
- Local bounds which are functions of the current notional: $R_n \in [l_n^{lo}(R_{n-1}), l_n^{hi}(R_{n-1})]$.
- Bounds that are functions of market data x_n (e.g. Libor or swap rates): $R_n \in [m_n^{lo}(x_n), m_n^{hi}(x_n)]$.

Denote by $\mathbb{C}_n(R_{n-1}, x_n)$ the set of time-T_n constraints, so that $R_n \in \mathcal{C}_n(R_{n-1}, x_n)$. The valuation of a flexi-swap can be done via a backward recursion equation: if $V_n(t, R)$ denotes the time-t value of the part of a flexi-swap paying strictly after T_n, given that $R_n = R$, we have

$$V_{n-1}(T_{n-1}, R) = P(T_{n-1}, T_n) X_{n-1} R + B(T_{n-1}) E_{T_{n-1}} \left[\frac{1}{B(T_n)} \max_{R' \in \mathcal{C}_n(R, x_n)} \{V_n(T_n, R')\} \right] \tag{139}$$

for $n = N, \ldots, 1$ and with terminal condition $V_N(T_N, R) \equiv 0$.

11.3.1 Purely local bounds

Assume that only local constraints are enforced and that these are of the form

$$\ell_n^{lo}(R_{n-1}) = \lambda_n^{lo} R_{n-1}, \quad \ell_n^{hi}(R_{n-1}) = \lambda_n^{hi} R_{n-1}, \quad 0 \leq \lambda_n^{lo} \leq \lambda_n^{hi}$$

Also assume for simplicity that the value of the flexi-swap scales linearly in notional, namely

$$V_n(T_n, R) = R V_n(T_n, 1)$$

Simple computations then show that the recursive equation (139) simplifies to

$$\begin{aligned} V_{n-1}(T_{n-1}) = {} & P(T_{n-1}, T_n) X_{n-1} \\ & + B(T_{n-1}) E_{T_{n-1}} \left[\frac{1}{B(T_n)} V_n(T_n) \left(\lambda_n^{lo} 1_{V_n(T_n)<0} + \lambda_n^{hi} 1_{V_n(T_n)>0} \right) \right] \end{aligned} \tag{140}$$

11.4 Fast pricing via exercise premia representation

On occasions the speed of valuation of Bermudan swaptions is more important than accuracy (e.g. robust hedging, calculation of CVA). We now consider a useful approximation based on the representation of a Bermudan swaption as a stream of coupons paid in the exercise region.

12 TARN's, volatility swaps and other derivatives

12.1 TARN's (targeted redemption notes)

A *TARN* pays structured coupons in exchange for Libor coupons until the cumulative amount of structured coupon payments exceeds a pre-agreed target R, at which point the derivative terminates. For a given tenor structure $0 = T_0 < T_1 < \cdots < T_N$, the time-0 value of a TARN is hence

$$\begin{aligned} V_{tarn}(0) &= E^B\left[\sum_{n=1}^{N-1} \frac{1}{B(T_{n+1})}\tau_n\left(C_n - L_n(T_n)\right)\mathbb{I}_{Q_n<B}\right], \\ Q_n &= \sum_{i=1}^{n-1}\tau_i C_i \end{aligned}$$

- Faithful reproduction of volatility smiles of various Libor rates is important for TARN's, so it is recommendable to use globally calibrated models with stochastic volatility.
- The knockout feature of the TARN will introduce *digital* discontinuities (jumps coming from the step function $\mathbb{I}_{Q_n<B}$), so Monte Carlo errors of the contract value and, especially, its risk sensitivities can be large. Ways of amending this are described in [AP10-3, Chapters 23 and 24].
- The full power of LM and gG models may not be required for TARN's can often do with a *local projection method*, as described below: in a nutshell, the idea is to find a simple local model that is calibrated the parts of a global model's volatility structure (e.g. LM) that are relevant to the derivative being valued, in such a way as to approximate the value of the global model for a particular derivative.

12.1.1 Local projection method

Consider the example of an inverse floating coupon $C_n = (s - g \cdot L_n(T_n))^+$. The time-0 value of the corresponding TARN with target B is

$$V_{tarn}(0) = E^B\left[\sum_{n=1}^{N-1} \frac{1}{B(T_{n+1})}\tau_n\left((s - g \cdot L_n(T_n))^+ - L_n(T_n)\right)\mathbb{I}_{\{\sum_{i=1}^{n-1}\tau_i(s-g\cdot L_i(T_i))^+<B\}}\right]$$

which depends on the values

$$\tilde{L} = (L_1(T_1), \ldots, L_{N-1}(T_{N-1}))$$

of Libor rates on their fixing dates only (the values at intermediate dates being irrelevant) so only the properties of the $(N-1)$-dimensional vector $\tilde{L}$ need be captured.

Assuming log-normal distributions for market rates, if two models assign the same values to the term variances of Libor rates $\mathrm{Var}(\ln L_n(T_n))$ and inter-temporal correlations of Libor rates $\mathrm{Corr}(\ln L_n(T_n))$, then the values of a TARN in both models should be the same. One should thus apply the local projection method as follows:

1. Calibrate a Libor market model to the full swaption volatility grid.
2. Use the calibrated LM model to calculate the relevant term volatilities and inter-temporal correlations needed for the TARN.
3. Pick a simpler model and calibrate it to the volatilities and correlations extracted from the LM model.
4. Use the calibrated local model for valuing the TARN and computing risk sensitivities.

12.1.2 Effect of volatility smiles

A successful candidate for the local model should have the ability to calibrate to volatility smiles of all Libor rates, in addition to having a low number of state variables, for instance the one-factor quasi-Gaussian model with stochastic volatility. One may proceed as follows:

1. Fix the mean reversion function to match the inter-temporal correlations of Libor rates $\mathrm{Corr}(\ln L_n(T_n), \ln L_m(T_m))$, $n, m = 1, \ldots, N-1$ as described in Section 6.2.4.
2. Calibrate to the volatility smiles of all Libor rates that appear in the payoff formula.
3. Choose the time-dependent local volatility function $\sigma_r(t, x, y)$ and volatility of variance function $\eta(t)$ to match the implied SV parameters of relevant caplets as in Section 6.2.2.

12.2 Volatility swaps

Volatility derivatives are contingent claims whose underlying is the volatility of a financial observable, rather than a financial observable itself. An (interest rate) *volatility swap* is a contract the measures realized volatility of a given rate $X_n(t)$. The most common coupons (uncapped and capped versions) are

$$C_n = |X_{n+1}(T_{n+1}) - X_n(T_n)|, \qquad C_n = \min\{|X_{n+1}(T_{n+1}) - X_n(T_n)|, c\}$$

The value of the structured leg of the volatility swap measures the realized variation of the rate $X_n(\cdot)$

$$\begin{aligned} V_{volswap}(t) &= \beta(t) E_t \left[\sum_{n=0}^{N-1} \tau_n \frac{1}{\beta(T_{n+1})} |X_{n+1}(T_{n+1}) - X_n(T_n)| \right] - V_{floatleg}(t), \quad (141) \\ V_{floatleg}(t) &= \sum_{n=0}^{N-1} \tau_n P(t, T_{n+1}) L_n(t) = P(t, T_0) - P(t, T_N) \end{aligned}$$

Common choice for the rate X_n are

- Fixed-tenor volatility swap, involving a swap rate of the same tenor on each of the fixing dates $X_n(t) = S_{n,m}(t)$, with fixed m.
- Fixed-expiry volatility swap, involving a swap rate with fixed expiry and tenor $X_n(t) = S_{K,m}(t)$.
- Volatility swaps on CMS rates, which measure the variation of the spread of two rates $X_n(t) = S_{n,m_1}(t) - S_{n,m_2}(t)$.

As was the case for TARN's, the value of a volatility swap (141) depends on the values of the swap rates S_n on their fixing dates only, so we may apply the local projection method (instead of a full-blown globally calibrated market model).

12.2.1 Volatility swaps with shout option

Some volatility swaps carry the option to *shout*, namely to choose when the fixing date occurs for the purpose of calculating the coupon payoff. The payoff for the n-th coupon is hence

$$C_n = |X_{n+1}(\eta_n) - X_n(T_n)|$$

with $\eta_n \in [T_n, T_{n+1}]$ chosen by the investor. This coupon is paid on date T_{n+1}. At time T_n, the option looks like an American option with exercise value

$$P(\eta_n, T_{n+1}) |S_{n+1}(\eta_n) - S_n(T_n)|$$

- **For valuation purposes, the shout option can be ignored for uncapped coupons**. By Jensen's inequality we have

$$\begin{aligned} P(\eta_n, T_{n+1}) E_{\eta_n}^{T_{n+1}} \left[|S_{n+1}(T_{n+1}) - S_n(T_n)| \right] &\geq P(\eta_n, T_{n+1}) \left| E_{\eta_n}^{T_{n+1}} \left[S_{n+1}(T_{n+1}) \right] - S_n(T_n) \right| \\ &\simeq P(\eta_n, T_{n+1}) \left| S_{n+1}(\eta_n) - S_n(T_n) \right| \end{aligned}$$

whence the early-exercise value is negligible.

- Consider now the case of a capped coupon with a shout option

$$C'_n = \min\{|S_{n+1}(\eta_n) - S_n(T_n)|, c\}$$

It seems as though one requires an estimation of the optimal exercise rule via regression, but it turns out that the situation is much simpler, as the following result shows.

Proposition 12.1 (Broadie and Detemple). *The value of an American option on a capped straddle with payoff $C'_n = \min\{|S_{n+1}(\eta_n) - S_n(T_n)|, c\}$ is equal to the value of a straddle with a barrier, so that $E_{T_n}^{T_{n+1}}[C'_n] = E_{T_n}^{T_{n+1}}[C''_n]$ where*

$$C''_n = c \cdot \mathbb{I}_{\{\max_{t \in [T_n, T_{n+1}]} \{|S_{n+1}(t) - S_n(t)|\} \geq c\}} + |S_{n+1}(t) - S_n(t)| \cdot \mathbb{I}_{\{\max_{t \in [T_n, T_{n+1}]} \{|S_{n+1}(t) - S_n(t)|\} < c\}}$$

so that the optimal exercise strategy is known analytically: for the period $[T_n, T_{n+1}]$ one should exercise the shout option on the first time t when $S_{n+1}(t)$ hits either of the barriers $S_n(t) \pm c$.

Having replaced an American option with a barrier option, we can value capped volatility swaps in standard Monte Carlo.

12.2.2 Min-Max volatility swaps

The structured coupon for a min-max volatility swap is

$$C_n = M_n - m_m, \quad \text{where} \quad M_n = \max_{t \in [T_n, T_{n+1}]} \{X_n(t)\}, \quad m_n = \min_{t \in [T_n, T_{n+1}]} \{X_n(t)\}$$

12.3 Forward swaption straddles

13 Risk management

Consider the following notation:

- $\Theta_{mkt}(t)$: N_{mkt}-dimensional vector of observable market data at time t (swap and futures rates for yield curve construction, cap and swaption prices for volatility calibration.
- $\Theta_{prm}(t)$: N_{prm}-dimensional vector of additional parameters that are not directly observed, but are estimated from historical data, such as short rate mean reversion, correlation parameterizations, SV mean reversion speeds, local volatility parameters, etc.
- $\Theta_{num}(t)$: additional parameters that control the numerical schemes used in the model, such as the number of MC paths, the size of the discretization steps, etc.

One first uses $\Theta_{mkt}(t)$ and $\Theta_{parm}(t)$ to construct the yield curve and calibrate the model, that is to obtain a new vector

$$\Theta_{mdl} = C(\Theta_{mkt}(t); \Theta_{prm}(t))$$

Given the time-t yield curve, a set of model parameters $\Theta_{mdl}(t)$ and a set of numerical scheme parameters $\Theta_{num}(t)$ we can compute the value $V(t)$ of our portfolio of derivative securities $V_i(t)$, namely

$$V(t) = V_1(t) + \cdots + V_n(t) = M(\Theta_{mdl}(t); \Theta_{num}(t)) = H(\Theta_{mkt}(t); \Theta_{prm}(t); \Theta_{num}(t)) \quad (142)$$

The role of a trading desk is then to construct a hedge around (142), with the hedge aiming to neutralize, in a standard Taylor-series sense, as many of the movements in the entire market data vector $\Theta_{mkt}(t)$ as possible.

The dimension of the market data vector N_{mkt} can be very high and it may be too costly to hedge against all components of $\Theta_{mkt}(t)$, so some type of principal component analysis may be undertaken.

13.1 Value at risk

The risk management team in a bank is primarily focused on analyzing the distribution of future portfolio values, in order to gauge the overall riskiness of the portfolio. The *value at risk* is one of the most commonly used risk measures.

VaR at level α, denoted Λ_α and typically a negative value, is the (1α)-percentile of the distribution of the $P\&L$ move $V(t+h) - V(t)$ in the real-life measure P, namely

$$P\left(V(t+h) - V(t) \leq \Lambda_\alpha | \mathcal{F}_t\right) = 1 - \alpha$$

In other words, the probability of losing more than $-\Lambda_\alpha$ over the time interval $[t, t+h]$ is less than $1 - \alpha$. Typically, α is set to 99% or 95%, and h to one business day.

Another commonly used risk measure is *conditional value-at-risk* (cVaR) which is defined as the conditional expectation

$$\Xi_{\alpha,h} = E^P\left[V(t+h) - V(t) | V(t+h) - V(t) \leq \Lambda_\alpha\right]$$

To compute Λ_α one needs a statistical description for the market data increment vector[41] δ, for which one may follow different methodologies:

[41]Namely, a vector of variables affecting the portfolio value, such as exchange rates, equity prices, interest rates and so forth.

1. Using the historical distribution of δ directly: ones takes the actual realizations of δ over the last N_{VaR} trading days and applies them to the current market data, thereby generating the empirical distribution of $V(t+h) - V(t)$.
   The calculation of VaR then amounts to ranking the impact of the last N_{VaR} market moves on the current portfolio, worst to best, and using the impact on the day with rank $(1-\alpha)N_{VaR}$ as VaR.
2. Using a parameterized, rather than historical, distribution of market moves. One assumes that
$$\delta_i \sim \mathcal{N}(0, s_i^2), \quad s_i = \sigma_i\sqrt{h}, \quad i = 1, \ldots, N_{mkt}$$
   where σ_i is annualized basis point volatility of the i-th market element.
   The co-dependence between the elements of δ is captured in a correlation matrix which is typically estimated from historical time series.
   To compute VaR in the Gaussian setup, we may use equation
$$V(t+h) = V(t) + \nabla^H(t)\delta + \frac{1}{2}\delta^\top A^H(t)\delta$$
   Computations are then analytically tractable and are encapsulated in the following proposition.

   Proposition 13.1. *Neglecting the Hessian $A^H(t) = 0$, namely assuming that*
$$V(t+h) = V(t) + \nabla^H(t)\delta$$
   and assuming that the elements of the N_{mkt}-dimensional vector δ have correlation matrix R and satisfy $\delta_i \sim \mathcal{N}(0, s_i^2)$ we have
$$\begin{aligned} \Lambda_\alpha &= v\Phi^{-1}(1-\alpha) \\ \Xi_\alpha &= -\frac{v}{1-\alpha}\phi\left(\Phi^{-1}(1-\alpha)\right), \\ v^2 &= \nabla^H(t)\,diag(s)R\,diag(s)\nabla^H(t)^\top \end{aligned}$$
   where $s = (s_1, \ldots, s_{N_{mkt}})$ and $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-x^2/2}\,dx$ is the Gaussian CDF.

   Proof. Note that the covariance matrix of δ is $C = \mathrm{diag}(s)R\mathrm{diag}(s)^\top$. Since $V(t+h) = V(t) + \nabla^H(t)\delta$, this implies that $V(t+h) - V(t) \sim N(0, v^2) \sim vN(0,1)$ where
$$v^2 = \nabla^H(t)C\nabla^H(t)^\top.$$
   Hence if $Z \sim \mathcal{N}(0,1)$ we have
$$P(V(t+h) - V(t) \le \Lambda_\alpha) = P\left(Z \le \frac{\Lambda_\alpha}{v}\right) = \Phi\left(\frac{\Lambda_\alpha}{v}\right) \equiv 1-\alpha \Longrightarrow \Lambda_\alpha = v\Phi^{-1}(1-\alpha)$$
   As for cVaR we have
$$E^P\left[\Delta V | \Delta V \le \Lambda_\alpha\right] = \frac{1}{P(\Delta V \le \Lambda_\alpha)}E^P\left[\Delta V \mathbb{I}_{\{\Delta V \le \Lambda_\alpha\}}\right] = \frac{v}{1-\alpha}E^P\left[Z\mathbb{I}_{\{Z \le \frac{\Lambda_\alpha}{v}\}}\right]$$
   □

Remark 13.1. 1. If we wish to include A^H in the computation, the distribution of $V(t+h) - V(t)$ is no longer simple. One can show that it is expressible as a sum of independent non-central chi-square random variables. From this representation, its characteristic function can be computed, and this can be turned numerically into a CDF from which VaR can be computed.
2. The key to a good VaR computation is a reliable and accurate estimate for ∇^H and A^H.

14 Payoff smoothing for Monte Carlo

- Practical risk management of a portfolio of interest rate securities revolves around *price sensitivities with respect to various valuation inputs*, such as market prices and model parameters. These sensitivities are computed by applying small perturbations to market and model parameters, followed by re-pricing of the securities portfolio in question.
- Price sensitivities are inherently less smooth than the price function itself, and this lack of smoothness will often put significant stress on a numerical scheme.
- It is thus important to try to adapt numerical schemes to avoid introducing spurious instabilities into the calculation of greeks.

We focus here on the so-called Tube Monte Carlo method for illustrative purposes[42] and we apply it to digital options, barrier options and CLE's. The method is designed to be applied when calculating sensitivities by direct perturbation.

14.1 Tube Monte Carlo for digital options

Consider a digital option with payoff

$$V_T = \mathbb{I}_{\{S(T)>B\}}$$

where $S(t)$ is the process for the underlying that is simulated using Monte Carlo. The standard Monte Carlo estimate (where $\{\omega_j\}$, $j = 1, \dots, J$ are the sample paths) is given by

$$V \simeq \frac{1}{J}\sum_{j=1}^{J} \mathbb{I}_{\{S(T,\omega_j)>B\}}$$

The idea is to replace the *point* sample of the payoff $\mathbb{I}_{\{S(T)>B\}}$ at $S(T,\omega_j)$ with an *average* estimate of the payoff in a small interval around $S(T,\omega_j)$, namely

$$\begin{aligned} V &\simeq \frac{1}{J}\sum_{j=1}^{J} V_j, \qquad V_j = E\left[\mathbb{I}_{\{S(T)>B\}}|A_j\right] \\ A_j &= \{\omega : S(T,\omega) \in [S(T,\omega_j)-\epsilon, S(T,\omega_j)+\epsilon], \quad \epsilon > 0\} \end{aligned}$$

Clearly:

- If $B \notin A_j$, then $E\left[\mathbb{I}_{\{S(T)>B\}}|A_j\right] = \mathbb{I}_{\{S(T,\omega_j)>B\}}$.
- If $B \in A_j$ and ϵ is small, using the uniform distribution approximation we obtain

$$\begin{aligned} E\left[\mathbb{I}_{\{S(T)>B\}}|A_j\right] &= Q\left(S(T) > B | S(T,\omega_j) - \epsilon \le S(T) < S(T,\omega_j)+\epsilon\right) \\ &= \frac{S(T,\omega_j)+\epsilon - B}{2\epsilon} \end{aligned}$$

Hence,

$$\begin{aligned} V_j &= E\left[\mathbb{I}_{\{S(T)>B\}}|A_j\right] \\ &= \frac{S(T,\omega_j)+\epsilon-B}{2\epsilon}\mathbb{I}_{\{S(T,\omega_j)-\epsilon\le B<S(T,\omega_j)+\epsilon\}} + 1\cdot\mathbb{I}_{\{S(T,\omega_j)+\epsilon\le B\}} + 0\cdot\mathbb{I}_{\{B<S(T,\omega_j)+\epsilon\}} \end{aligned}$$

[42] For a detailed discussion, including payoff smoothing methods for PDE's, see [AP10-3, Chapter 23].

14.2 Tube Monte Carlo for barrier options

Consider a derivative which is a knock-in barrier into a stream of coupons $X_1, \ldots, X_{N-1}$, with the knock-in feature defined by a stopping time index η, so that the derivative pays coupons X_i at T_{i+1} for $i = \eta, \ldots, N-1$ and has value

$$V(0) = E\left[\sum_{i=1}^{N-1} \frac{1}{B(T_{i+1})} X_i \mathbb{I}_{\{i \geq \eta\}}\right]$$

The standard Monte Carlo estimate is

$$V(0) = \frac{1}{J}\sum_{j=1}^{J}\left(\sum_{i=1}^{N-1} \frac{1}{B(T_{i+1}, \omega_j)} X_i(\omega_j) \mathbb{I}_{\{i \geq \eta(\omega_j)\}}\right)$$

As above, the indicator $\mathbb{I}_{\{i \geq \eta(\omega_j)\}}$ will introduce digital discontinuities in the payoff which lead to poor stability and risk sensitivities. The idea of the tube method is again to replace the point estimates of the payoff with payoff averages over appropriately defined tubes. If $x(t, \omega)$ denotes the d-dimensional vector of state variables of the underlying model, define

$$\begin{aligned} A_j^\epsilon &= \bigcap_{i=1}^{N-1} A_{j,i}^\epsilon, \\ A_{j,i}^\epsilon &= \{\omega : \|x(T_i, \omega) - x(T_i, \omega_j)\| < \epsilon\} \end{aligned}$$

Observe that A_j^ϵ denotes the set of sample paths that come within ϵ-distance of $x(T_i, \omega_j)$ for all T_i, $i = 1, \ldots, N-1$. We may then take our new Monte Carlo estimator to be

$$V(0) = \frac{1}{J}\sum_{j=1}^{J} V_j, \quad V_j = E\left[\sum_{i=1}^{N-1} \frac{1}{B(T_{i+1}, \omega)} X_i(\omega) \mathbb{I}_{\{i \geq \eta(\omega)\}} | A_j^\epsilon\right]$$

Being smooth functions of the path ω, we may just evaluate $\frac{1}{B(T_{i+1},\omega)}$ and $X_i(\omega)$ at the sample path, so that V_j becomes

$$V_j = \sum_{i=1}^{N-1} \frac{1}{B(T_{i+1}, \omega_j)} X_i(\omega_j) E\left[\mathbb{I}_{\{i \geq \eta(\omega)\}} | A_j^\epsilon\right]$$

An approximation of the probabilities $q_i(\omega_j) := E\left[\mathbb{I}_{\{i \geq \eta(\omega)\}} | A_j^\epsilon\right]$ is provided in [AP10-3, Proposition 23.4.1] under the assumption that stopping time index η can be written as the first hitting time of a state-dependent boundary

$$\eta(\omega) = \min\{n \geq 1 : \psi_n(x(T_n, \omega)) \geq 0\} \cap N$$

and is given by

$$\begin{aligned} 1 - q_i(\omega_j) &= \prod_{n=1}^{i}(1 - p_n(\omega_j)), \\ p_n(\omega_j) &= \begin{cases} 1, & \psi_{n,j} - \delta_{n,j} \geq 0, \\ \frac{\psi_{n,j} + \delta_{n,j}}{2\delta_{n,j}}, & \psi_{n,j} - \delta_{n,j} < 0 < \psi_{n,j} + \delta_{n,j} \\ 0, & \psi_{n,j} + \delta_{n,j} \leq 0 \end{cases} \end{aligned}$$

where

$$\psi_{n,j} = \psi_n(x(T_n, \omega_j)), \qquad \delta_{n,j} = \epsilon\|\nabla\psi_{n,j}\|, \qquad \|\nabla\psi_{n,j}\| = \nabla\psi_n(x)_{|x=x(T_n,\omega_j)}$$

Remark 14.1 (Tube Monte Carlo for CLE's)**.** Recall that CLE's can be valued in Monte Carlo by representing them as knock-in discrete-barrier options with the knock-in defined by an estimate of the exercise index[43] so we may apply the tube Monte Carlo method we just outlined to smooth out the payoff. Concretely, the function $\psi(x(T_n, \omega))$ in this case would be

$$\psi(x(T_n, \omega)) = \mathcal{C}(\tilde{U}_n(T_n))^\top \zeta(T_n, \omega_k') - \mathcal{C}(\tilde{H}_n(T_n))^\top \zeta(T_n, \omega_k')$$

Remark 14.2 (Tube Monte Carlo for TARN's)**.** A TARN can be represented as a derivative that pays a stream of coupons until a knock-out event takes place when a sum of structured coupons exceeds a certain target.

$$\begin{aligned} V_{TARN}(0) &= E\left(\sum_{i=1}^{N-1} \frac{1}{B(T_{i+1})} X_i \mathbb{I}_{\{i \leq \eta\}}\right), \\ \eta(\omega) &= \min\{n \geq 1 : Q_n(\omega) - R \geq 0\} \wedge N \end{aligned}$$

[43]C.f. equation (136).

15 Pathwise differentiation

We start by recalling the pathwise differentiation for European options, following [AP10-1, Section 3.3.2].

Take a European T-maturity option on a dividend-free stock with price process $S(t)$ with payout function $g(S(T))$ and consider the problem of computing

$$\frac{dV}{dS_0} = \lim_{h \to 0} \frac{V(S_0 + h) - V(S_0)}{h}$$

By the risk-neutral pricing formula $V(S_0) = E\left[g(S(T))\right]$. Assuming that the stock price follows a GBM we have

$$\begin{aligned} S(T) &= S_0 \exp\left[\left(r - \frac{\sigma^2}{2}\right) T + \sigma \sqrt{T} Z\right], \qquad Z \sim N(0,1) \\ S_h(T) &= (S_0 + h) \exp\left[\left(r - \frac{\sigma^2}{2}\right) T + \sigma \sqrt{T} Z\right] \end{aligned}$$

so $S_h(T) = (S_0 + h)\frac{S(T)}{S_0}$. If the payout function $g(\cdot)$ is regular enough so that the expectation and the limit are interchangeable, then

$$\frac{dV}{dS_0} = e^{-rT} E\left[\lim_{h \to 0} \frac{g(S_h(T)) - g(S(T))}{h}\right] = e^{-rT} E\left[g'(S(T)) \frac{S(T)}{S_0}\right] \tag{143}$$

so we can implement (143) by generating samples of $S(T)$ and recording the sample averages of $g'(S(T))\frac{S(T)}{S_0}$.

Remark 15.1. For discontinuous payoffs, such as $g(x) = 1_{\{x>K\}}$ care must be taken upon implementing this. In this case, (143) reads

$$\frac{dV}{dS_0} = e^{-rT} E\left[\delta(S(T) - K) \frac{S(T)}{S_0}\right] = e^{-rT} \frac{K}{S_0} \varphi_S(K)$$

where $\delta(\cdot)$ is the delta function. This is correct (provided that the terminal stock density $\varphi_S(\cdot)$ is known, but it is unsuited for Monte Carlo simulation since the likelihood of $\delta(S(T) - K)$ being non-zero is zero.

More generally, consider estimating $\frac{dV}{d\alpha}$ where $V(\alpha) = E[Y(\alpha)]$. Under certain regularity conditions (c.f. [AP10-1, Proposition 3.3.1]), we seek to write

$$\frac{dV}{d\alpha} = \frac{d}{d\alpha} E[Y(\alpha)] = E\left[\frac{d}{d\alpha} Y(\alpha)\right]$$

Assuming that Y represents a payout function, namely

$$Y(\alpha) = g(X(\alpha)), \quad X(\alpha) = (X_1(\alpha, \ldots, X_q(\alpha))$$

we can further write

$$\frac{d}{d\alpha} Y(\alpha) = \sum_{i=1}^{q} \frac{\partial g}{\partial X_i} \frac{dX_i}{d\alpha}$$

In order to compute the derivatives $\frac{dX_i}{d\alpha}$, assuming that $X(t)$ is driven by a scalar SDE

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

differentiating with respect to α on both sides and writing $D_\alpha(t) = \frac{dX(t)}{d\alpha}$ we obtain

$$dD_\alpha(t) = \frac{d}{dx}\mu(t, X(t))D_\alpha(t)dt + \frac{d}{dx}\sigma(t, X(t))D_\alpha(t)dW(t)$$

and the latter SDE can be discretized and simulated in parallel with the simulation of the SDE for $X(t)$ itself.

We next illustrate how to apply this method to compute greeks of CLE's and barrier options, but we first make a few comments on the likelihood ratio method.

15.1 On the unsuitability of the likelihood ratio method in interest rate modeling

In general, an alternative method to compute sensitivities by Monte Carlo simulation is the likelihood ratio method. In general, assume that $Y(\alpha)$ represents a deflated payout function $g(\cdot)$ applied to a vector of random variables $X(\alpha) = (S_1(\alpha), \quad S_q(\alpha))$ with joint density $f(x;\alpha)$, $x \in \mathbb{R}^q$. Then

$$V(\alpha) = E\left[g(S(T))\right] = \int_{\mathbb{R}^q} g(x)f(x;\alpha)\,dx$$

Assuming that the density function $f(x;\alpha)$ is a smooth, we can exchange differentiation and integration

$$\frac{\partial V(\alpha)}{\partial \alpha} = \int_{\mathbb{R}^q} g(x)\frac{\partial f(x;\alpha)}{\partial \alpha}\,dx = \int_{\mathbb{R}^q} g(x)\frac{\partial \ln f(x;\alpha)}{\partial \alpha} f(x;\alpha)\,dx = E\left[g(X(T))\ell(X(T))\right]$$

where $\ell(x)\frac{\partial \ln f(x;\alpha)}{\partial \alpha}$ is the so-called *log-likelihood ratio.*

Example 15.1. In the Black-Scholes setting $S(T) = S_0 \exp\left[\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z\right]$, $Z \sim \mathcal{N}(0,1)$ and $Y(T) = \ln S(T)$, the log-likelihood ratio is given by

$$\ell(Y(T)) = \frac{\partial \ln \varphi(Y(T); S_0)}{\partial S_0} = \frac{1}{S_0\sigma\sqrt{T}}Z$$

- The likelihood ratio method does not require smoothness of the payout function, but only of the density function, which is a mild assumption.
- Knowledge of the density function is required, although one always has the option of using a Gaussian approximation based on an Euler discretization.
- In general, it is the earliest observation T_1 of the underlying asset process[44] that determines how fast its variance explodes (note the $\sqrt{T}$ in the denominator of the log-likelihood ratio in the Black-Scholes model). Because of the regular structure of most interest rate derivatives, the time T_1 will in most cases be rather short, resulting in high variance of the estimate.

[44]For instance, the first coupon fixing date, the first exercise date of a CLE or the first knock-out date of a barrier.

15.2 CLE's

Recall the main valuation recursion for CLE's (in the spot measure Q^B

$$
\begin{aligned}
\frac{H_{n-1}(T_{n-1})}{B(T_{n-1})} &= E^B_{T_{n-1}}\left[\frac{1}{B(T_n)}\max\{H_n(T_n), U_n(T_n)\}\right], \\
U_n(t) &= B(t)\sum_{i=n}^{N-1} E_t\left[\frac{1}{B(T_{i+1})}X_i\right], \\
X_i &= \tau_i(C_i - L_i(T_i))
\end{aligned}
$$

where X_i are the net coupon payments, C_i are structured coupons, $H_n(t)$ is the n-th hold value and $U_n(t)$ is the n-th exercise value. Denote by Δ_α a pathwise differentiation operator with respect to a given parameter α.

Proposition 15.1. *Assuming that the coupons X_n, $n = 1, \ldots, N-1$ and the inverse numraire $\frac{1}{B(t)}$ are Lipschitz continuous functions of the parameter α, for any $n = 1, \ldots, N-1$ we have*

$$
\begin{aligned}
\Delta_\alpha\left(\frac{H_{n-1}(T_{n-1})}{B(T_{n-1})}\right) = E_{T_{n-1}}&\left[\Delta_\alpha\left(\frac{U_n(T_n)}{B(T_n)}\right)\mathbb{I}_{\{U_n(T_n)>H_n(T_n)\}}\right] \\
&+ E_{T_{n-1}}\left[\Delta_\alpha\left(\frac{H_n(T_n)}{B(T_n)}\right)\mathbb{I}_{\{H_n(T_n)>U_n(T_n)\}}\right]
\end{aligned}
$$

Unwrapping this recursive statement, if η is the optimal exercise time index then

$$
\Delta_\alpha(H_0(0)) = E\left[\sum_{n=\eta}^{N-1}\Delta_\alpha\left(\frac{X_n}{B(T_{n+1})}\right)\right]
$$

Remark 15.2. From the fact that

$$
H_0(0) = E\left[\sum_{n=\eta}^{N-1}\frac{X_n}{B(T_{n+1})}\right], \quad \Delta_\alpha(H_0(0)) = E\left[\sum_{n=\eta}^{N-1}\Delta_\alpha\left(\frac{X_n}{B(T_{n+1})}\right)\right]
$$

is looks as though one can compute the derivative Δ_α by differentiating the first sum and pretending that the optimal exercise time index η is independent of α. This is not the case in most cases: see [AP10-3, Section 24.1.1.2] for a justification.

Remark 15.3. When using brute-force perturbation methods to evaluate CLE greeks, there are different alternatives when deciding how to treat the exercise decision in the perturbed market data scenario. As justified in [AP10-3, Section 24.1.1.3], it is convenient to re-use the estimate of the optimal exercise index from the base scenario in order to avoid introducing discontinuities.

15.3 Barrier options

Recall that a CLE can be interpreted as a knock-in barrier option (where the barrier condition is defined by the optimal exercise rule) which suggests that the pathwise differentiation method outlined in the previous section could be extended to general barrier options.

15.3.1 Digital option warm-up

Consider a T-maturity European option with payoff

$$X = \mathbb{I}_{\{G>h\}}R, \qquad G, R \quad \mathcal{F}_T\text{-measurable}, \qquad h \in \mathbb{R}$$

Differentiating with respect to α we obtain

$$\Delta_\alpha \left(\mathbb{I}_{\{G>h\}}R\right) = \mathbb{I}_{\{G>h\}}\Delta_\alpha R + \frac{\partial \mathbb{I}_{\{G>h\}}}{\partial G} R\Delta_\alpha G = \mathbb{I}_{\{G>h\}}\Delta_\alpha R + \delta(G-h)R\Delta_\alpha G$$

Hence,

$$\begin{aligned}
\Delta_\alpha E\left[\frac{X}{B(T)}\right] &\overset{[*]}{=} E\left[\Delta_\alpha\left(\frac{X}{B(T)}\right)\right] \\
&= E\left[\Delta_\alpha\left(\frac{1}{B(T)}\right)\mathbb{I}_{\{G>h\}}R\right] + E\left[\frac{1}{B(T)}\mathbb{I}_{\{G>h\}}\Delta_\alpha R\right] + E\left[\frac{1}{B(T)}\delta(G-h)R\Delta_\alpha G\right] \\
&= E\left[\Delta_\alpha\left(\frac{1}{B(T)}\right)\mathbb{I}_{\{G>h\}}R\right] + E\left[\frac{1}{B(T)}\mathbb{I}_{\{G>h\}}\Delta_\alpha R\right] \\
&\quad + \gamma_G(h)\cdot E\left[\frac{1}{B(T)}R\Delta_\alpha G | G = h\right]
\end{aligned}$$

where $\gamma_G(h)$ is the PDF of G evaluated at h.

- In $[*]$ we assumed that we can exchange Δ_α and $E[\cdot]$, even though we can't due to the digital discontinuity of $\mathbb{I}_{\{G>h\}}$. The result of the computation is nevertheless true, and can be justified using *Malliavin Calculus*.
- The expected values on the RHS can be computed in a numerical scheme such as Monte Carlo as long as the density $\gamma_H(\cdot)$ is known and the conditional expectation

  $$E\left[\frac{1}{B(T)}R\Delta_\alpha G | G = h\right]$$

  can be evaluated. Both can be computed using Malliavin calculus, even though in some cases these can be approximated in closed form.

15.3.2 Barrier options

Consider a barrier schedule $\{T_n\}_{n=1}^{N-1}$ to which we associate knockout variables G_n and barrier levels h_n. Consider an option that pays R_n on the first date T_n where $G_n > h_n$, so that

$$\begin{aligned}
V(0) &= E\left[\frac{R_\eta}{B(T_\eta)}\right], \\
\eta &= \min\{k \geq 1 : G_k > h_k\} \wedge N
\end{aligned}$$

More generally, consider the option with the barrier condition checked at times $T_{n+1}, \ldots, T_{N-1}$ only, namely

$$\begin{aligned}
V_n(t) &= B(t)E_t\left[\frac{R_{\eta_n}}{B(T_{\eta_n})}\right], \\
\eta_n &= \min\{k \geq n+1 : G_k > h_k\} \wedge N
\end{aligned}$$

Proposition 15.2. *(c.f. [AP10-3, Proposition 24.1.3]) For the barrier option described above, which pays R_n on the first T_n where $G_n > h_n$, $n = 1, \ldots, N-1$, the pathwise derivative with respect to a parameter α is given by*

$$\Delta_\alpha V(0) = E\left[\frac{1}{B(T_\eta)}\,\Delta_\alpha R_\eta|_{n=\eta}\right] + E\left[\frac{1}{B(T_\eta)}\,\gamma_\eta(h_\eta)\,(R_\eta - V_\eta(T_\eta))\,\Delta_\alpha G_n|_{n=\eta}\middle|\, G_\eta = h_\eta\right]$$

Sketch of proof. The following recursion is clear

$$\frac{V_{ki,n}(T_n)}{B(T_n)} = E_{T_n}\left[\frac{R_{n+1}}{B(T_{n+1})}\mathbb{I}_{\{G_{n+1}>h_{n+1}\}}\right] + E_{T_n}\left[\frac{1}{B(T_{n+1})}V_{ki,n+1}(T_{n+1})\mathbb{I}_{\{G_{n+1}\le h_{n+1}\}}\right]$$

Formal differentiation yields the following recursion for $\Delta_\alpha V_{ki,n}(T_n)$

$$\begin{aligned}\Delta_\alpha\left(\frac{V_{ki,n}(T_n)}{B(T_n)}\right) &= E_{T_n}\left[\Delta_\alpha\left(\frac{R_{n+1}}{B(T_{n+1})}\right)\mathbb{I}_{\{G_{n+1}>h_{n+1}\}}\right]\\ &\quad + E_{T_n}\left[\Delta_\alpha\left(\frac{1}{B(T_{n+1})}V_{ki,n+1}(T_{n+1})\right)\mathbb{I}_{\{G_{n+1}\le h_{n+1}\}}\right]\\ &\quad + E_{T_n}\left[\frac{1}{B(T_{n+1})}\,(R_{n+1} - V_{ki,n+1}(T_{n+1}))\,\delta(G_{n+1} - h_{n+1})\right]\end{aligned}$$

Unwrapping the recursion the claim follows. □

Remark 15.4. The method applies directly to CLE's through their interpretation as knock-in options. Recall that

$$V_{CLE}(0) = E\left[\frac{U_\eta(T_\eta)}{B(T_\eta)}\right] = E\left[\sum_{n=\eta}^{N-1}\frac{X_n}{B(T_{n+1})}\right], \quad \eta = \min\{n : U_n(T_n) > H_n(T_n)\}$$

so that Proposition 15.1 follows directly from Proposition 15.2 upon setting $R_n = U_n(T_n)$, $G_n = U_n(T_n) - H_n(T_n)$ and $h_n = 0$.

Remark 15.5. These results are rather complex and require knowledge of the transition densities and conditional probabilities, which are often hard to compute. Therefore, it may be more fruitful to apply payoff smoothing techniques (e.g. tube Monte Carlo method) or importance sampling techniques to integrate any discontinuities before applying the pathwise differentiation method.

15.4 Pathwise differentiation for Monte Carlo based models

Consider an LM model with separable deterministic local volatility with dynamics in the spot measure Q^B given by

$$\begin{aligned} dL_n(t) &= \varphi(L_n(t))\lambda_n(t)^\top\left(\mu_n(t)dt + dW^B(t)\right)\\ \mu_n(t) &= \sum_{j=q(t)}^{n}\frac{\tau_j\varphi(L_j(t))\lambda_j(t)}{1+\tau_j L_j(t)}\end{aligned}$$

Assume that the LM model and the security to be priced share the same tenor structure $0 = T_0 < \cdots < T_N$.

16 Importance sampling and control variates

In this section we explore two common variance reduction techniques (importance sampling and control variates). We start by outlining the basic principles upon which they rely and by illustrating them via simple examples extracted from [Gla03, Chapter 4]. We then describe applications to fixed-income derivatives following [AP10-3, Chapter 25].

16.1 Importance sampling

The idea of *importance sampling* is to use a change of measure to reduce variance. Assume we seek to estimate $\mu = E^P[Y]$. Let $\hat{P}$ be an equivalent measure to P; then

$$\mu = E^P[Y] = E^{\hat{P}}\left[Y\frac{dP}{d\hat{P}}\right] \stackrel{not}{=} E^{\hat{P}}\left[\frac{Y}{R}\right]$$

where $R = \frac{d\hat{P}}{dP}$ is the Radon-Nikodym derivative. The goal of the method is to choose the new measure $\hat{P}$ optimally in order to minimize the $\hat{P}$-variance of $\frac{Y}{R}$.

Assume $Y = g(X)$ where $g : \mathbb{R}^p \to R$ is a *well-behaved* function and X is a p-dimensional random vector with density $f_X : \mathbb{R}^p \to \mathbb{R}$. Then

$$\mu = E^P[g(X)] = \int_{\mathbb{R}^p} g(x) f_X(x)\, dx$$

can be estimated by $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^n g(X_i)$ for independent samples $X_1, \ldots, X_n$ of X.

If $h : \mathbb{R}^p \to \mathbb{R}$ is the (strictly positive) density function of X in a nother measure $\hat{P}$, we can represent μ as

$$\mu = E^P[g(X)] = \int_{\mathbb{R}^p} g(x)\frac{f_X(x)}{h(x)} h(x)\, dx = E^{\hat{P}}\left[g(X)\frac{f(X)}{h(X)}\right]$$

which in turn has a Monte Carlo estimate $\bar{\mu}_n^h = \frac{1}{n}\sum_{i=1}^n g(X_i)\frac{f(X_i)}{h(X_i)}$, where now $X_1, \ldots, X_n$ are independent draws from h. The quotient $\frac{f(x)}{(h(x)}$ is the so-called *likelihood ratio.*

We now compare both variances:

$$\begin{aligned}
\mathrm{Var}[\bar{\mu}_n^h] &= \frac{1}{n}\left(E^{\hat{P}}\left[g(X)^2\frac{f(X)^2}{h(X)^2}\right] - \mu^2\right) = \frac{1}{n}\left(E^P\left[g(X)^2\frac{f(X)}{h(X)}\right] - \mu^2\right)\\
\mathrm{Var}[\bar{\mu}_n] &= \frac{1}{n}\left(E^P\left[g(X)^2\right] - \mu^2\right)
\end{aligned}$$

and we note that importance sampling will lower variance if

$$E^P\left[g(X)^2\frac{f(X)}{h(X)}\right] < E^P\left[g(X)^2\right]$$

Now assume that the p-dimensional process $X(t)$ is driven by an SDE

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dW(t) \tag{144}$$

where $W(t)$ is a d-dimensional Brownian motion and that we seek to evaluate

$$E^P[g(X(T))]$$

Assume that the SDE is simulated by an m-dimensional scheme so that

$$g(X(T)) = G(Z_1, \ldots, Z_m), \quad G : \mathbb{R}^{p\times m} \to \mathbb{R}$$

where $Z_1, \dots, Z_m$ are independent p-dimensional Gaussian vectors with density $\phi(z)$. By independence,

$$E^P[g(X(T))] = \int_{\mathbb{R}^{p\times m}} G(z_1, \dots, z_m) \prod_{i=1}^{m} \phi(z_i)\, dz_i$$

The likelihood ratio $\ell(z) = \prod_i^m \frac{\phi(z_i)}{h(z_i)}$ performs a change of measure (to $\hat{P}$, say) that preserves independence of Z_i and alters the common marginal density from $\phi(z)$ to $h(z)$ so that

$$E^P[g(X(T))] = E^{\hat{P}}\left[G(Z_1, \dots, Z_m) \prod_{i=1}^{m} \frac{\phi(Z_i)}{h(Z_i)}\right]$$

Example 16.1. Assume $p = 1$ and consider shifting the means of the Z_i from 0 to some scalar μ, so that

$$h(z_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(z_i - \mu)^2\right)$$

whereby

$$\ell(z; \mu) = \exp\left(-\mu \sum_{i=1}^{m} z_i + \frac{m}{2}\mu^2\right)$$

The idea would then be to set μ to minimize the $\hat{P}$-variance of the term $G(Z)\ell(Z; \mu)$. This minimization problem is generally handled numerically.

Example 16.2. (c.f. [AP10-1, Section 3.4.4.5]) Consider the problem of estimating by Monte Carlo

$$P(Z > c)$$

where $Z \sim \mathcal{N}(0, 1)$ and large c, so that the event $\{Z > c\}$ is rare and a standard Monte Carlo simulation might be inaccurate.

In ordinary Monte Carlo we would use the sample mean estimator

$$P(Z > c) = E^P[\mathbb{I}_{\{Z>c\}}] \simeq \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{Z_i>c\}}$$

where $Z_1, \dots, Z_n \sim \mathcal{N}(0, 1)$ are independent standard Gaussian samples. The variance of this estimator is easily seen to be

$$\mathrm{Var}^P(I_{\{Z>c\}}) = P(Z > c)\,[1 - P(Z > c)]$$

Introducing a probability that shifts the mean of Z from 0 to some μ, by Example 16.1 the likelihood ratio is

$$\ell(z) = \frac{dP}{d\hat{P}} = e^{-\mu z + \frac{1}{2}z^2}$$

so

$$P(Z > c) = E^P[\mathbb{I}_{\{Z>c\}}] = E^{\hat{P}}\left[e^{-\mu z + \frac{1}{2}z^2}\mathbb{I}_{\{Z>c\}}\right], \quad Z \sim \mathcal{N}(\mu, 1)$$

and a Monte Carlo estimator for this is then

$$\frac{1}{n}\sum_{i=1}^{n} e^{-\mu(Z_i+\mu)+\frac{1}{2}z^2}\mathbb{I}_{\{Z_i+\mu>c\}}, \quad Z_1, \dots, Z_n \sim \mathcal{N}(0, 1)$$

As for the variance of the estimator one can compute that

$$\mathrm{Var}^{\hat{P}}\left(e^{-\mu z+\frac{1}{2}z^2}\mathbb{I}_{\{Z>c\}}\right) = e^{\mu^2} P(Z > \mu + c) - P(Z > c)^2$$

and the choice of μ that minimizes this $\hat{P}$-variance is

$$\mu^* = \arg\min_\mu \left\{e^{\mu^2} P(Z > \mu + c)\right\} = \{\mu^* : 2\mu^* \left[1 - \mathcal{N}(c+\mu^*)\right] - \mathfrak{n}(c+\mu^*)\}$$

This holds more generally in higher dimensions and for more general payoffs. Assume we want to estimate $E[G(\boldsymbol{Z})]$ and assume that we restrict ourselves to changes of distribution that change the mean of $\boldsymbol{Z}$ from $\mathbf{0}$ to $\boldsymbol{\mu}$. In this case we have

$$\begin{aligned}
E[G(\boldsymbol{Z})] &= E^{P_{\boldsymbol{\mu}}}\left[G(\boldsymbol{Z}) e^{-\boldsymbol{\mu}^\top Z + \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\mu}}\right], \quad \boldsymbol{\mu} \in \mathbb{R}^n, \quad \boldsymbol{Z} \overset{P_{\boldsymbol{\mu}}}{\sim} \mathcal{N}(\boldsymbol{\mu}, \mathbb{I}) \\
&\overset{[1]}{\simeq} E^{P_{\boldsymbol{\mu}}}\left[e^{F(\boldsymbol{Z})} e^{-\boldsymbol{\mu}^\top Z + \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\mu}}\right] \\
&\simeq E\left[e^{F(\boldsymbol{\mu}+\boldsymbol{Z})} e^{-\boldsymbol{\mu}^\top Z - \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\mu}}\right] \\
&\overset{[2]}{\simeq} E\left[e^{F(\boldsymbol{\mu}) + \nabla F(\boldsymbol{\mu})\boldsymbol{Z}} e^{-\boldsymbol{\mu}^\top Z - \frac{1}{2}\boldsymbol{\mu}^\top \boldsymbol{\mu}}\right]
\end{aligned}$$

where in [1] we expressed $F(\boldsymbol{Z}) = \ln G(\boldsymbol{Z})$ and in [2] we linearized $F(\boldsymbol{\mu} + \boldsymbol{Z})$. Note that choosing

$$\nabla F(\boldsymbol{\mu})\boldsymbol{Z} = \boldsymbol{\mu}^\top \tag{145}$$

we eliminate the dependence on $\boldsymbol{Z}$, and hence the variance of the estimator (although [2] is just an approximation, so we end up with a low-variance estimator).

Example 16.3. (c.f. [Gla03, Section 4.6.2]) Consider for instance the case of an arithmetic Asian option. Ignoring the discount factor e^{-rT} we have a payoff

$$\begin{aligned}
G(\boldsymbol{Z}) &= \left[\bar{S}(\boldsymbol{Z}) - K\right]^+, \quad \bar{S} = \frac{1}{m}\sum_{i=1}^m S(t_i), \quad 0 = t_0 < t_1 < \cdots < t_m \\
S(t_i) &= S(t_{i-1}) \exp\left((r - \frac{1}{2}\sigma^2)(t_i - t_{i-1}) + \sigma\sqrt{t_i - t_{i-1}} Z_i\right), \quad Z_i \sim \mathcal{N}(0,1)
\end{aligned} \tag{146}$$

Imposing condition (145)

$$\begin{aligned}
\nabla(\ln G(\boldsymbol{Z})) = \boldsymbol{Z} \quad &\Longrightarrow \quad \frac{\partial}{\partial Z_j}\ln(\bar{S}(\boldsymbol{Z}) - K) = \frac{1}{\bar{S} - K}\frac{\partial \bar{S}}{\partial Z_j} = Z_j \\
&\Longrightarrow \quad \frac{\partial \bar{S}}{\partial Z_j} - (\bar{S} - K) Z_j = 0
\end{aligned}$$

Using the discretization of S this condition becomes

$$Z_j = \frac{1}{mG(\boldsymbol{Z})}\sum_{i=j}^m \sigma\sqrt{t_i - t_{i-1}} S(t_i)$$

which, given an equally spaced time grid with $t_i - t_{i-1} = h$ yields a recursion

$$z_1 = \frac{\sigma\sqrt{h}[G(\boldsymbol{Z}) + K]}{G(\boldsymbol{Z})}, \quad z_{j+1} = z_j - \frac{\sigma\sqrt{h} S(t_j)}{mG(\boldsymbol{Z})}, \quad j = 1, \ldots, m-1 \tag{147}$$

Given a payoff vale $y \equiv G(\boldsymbol{Z})$, this scheme determines $\boldsymbol{Z}$ (if $y \equiv G(\boldsymbol{Z})$, apply (147) to calculate Z_1, ten apply (ch25-importance-sampling-example-1) to calculate $S(t_1)$, then apply (ch25-importance-sampling-example-2) again to obtain Z_2 and so forth). Each payoff value $y = G(\boldsymbol{Z})$ thus determines $\boldsymbol{Z}$, and hence a path $\{S(t_i; y)\}_{i=1}^m$ and we seek to find the y for which

$$\frac{1}{m}\sum_{i=1}^m S(t_i; y) - K = y$$

This equation can be solved for y_* easily using a 1-dimensional search and this determines $\boldsymbol{Z}_* = \boldsymbol{Z}(y_*)$. To simulate, we then simply set $\boldsymbol{\mu} = \boldsymbol{Z}_*$ and apply importance sampling with mean $\boldsymbol{\mu}$.

16.2 Payoff smoothing via importance sampling

We illustrate how use the importance sampling method to produce payoff smoothing by considering two examples: binary options and TARN's.

16.2.1 Binary options

Assume $X \sim \mathcal{N}(\mu, \sigma^2)$ and consider en option that pays $g(X)$ provided that X is below a certain barrier b, namely

$$V = E^P\left[g(X)\mathbb{I}_{\{X<b\}}\right], \quad P \text{ some pricing measure}$$

- Valuing by Monte Carlo entails simulating independent Gaussian samples, discarding those that end up above the barrier b and averaging the payoff values over non-discarded samples.
- If b is low, the proportion of samples that contribute to the average is small, which leads to large simulation error.
- The digital feature of the payoff reduces the accuracy and stability of Monte Carlo estimates of greeks.

We thus seek to change the probability to increase the proportion of *interesting* samples. Conditioning on the survival we obtain

$$V = E[g(X)|X<b]P(X<b) = E[g(X)|X<b]\mathcal{N}\left(\frac{b-\mu}{\sigma}\right) = E\left[g(\mathcal{N}^{-1}(U'))\right]\mathcal{N}\left(\frac{b-\mu}{\sigma}\right)$$

where $U' \sim U(0, \mathcal{N}(b))$. Since g is smooth, a Monte Carlo evaluation of $E\left[g(\mathcal{N}^{-1}(U'))\right]$ will have good convergence and exhibit stable greek estimates.

Setting $\Lambda = \frac{\mathbb{I}_{\{X<b\}}}{P(X<b)}$ we can write

$$V = E\left[g(X)\mathbb{I}_{\{X<b\}}\right] = E[g(X)\Lambda]P(X<b) = \tilde{E}[g(X)]P(X<b), \quad \Lambda = \frac{d\tilde{P}}{dP}$$

so conditioning on survival can be interpreted as a change of measure defined by the Radon-Nikodym derivative Λ.

16.2.2 TARN's

Recall that the TARN valuation formula under the spot measure Q^B reads

$$V_{tarn}(0) = E^B\left[\sum_{n=1}^{N-1}\frac{1}{B(T_{n+1})}X_n(T_n)\mathbb{I}_{\{Q_b<R\}}\right], \quad Q_n = \sum_{i=1}^{n-1}\tau_i C_i$$

where the X_n's are net coupons and R is the total return.

Example 16.4 (Removing first digital discontinuity)**.** We first try to apply the approach from the previous section to remove the first digital discontinuity. Assume structured coupons of the form

$$C_n = (s - gL_n(T_n))^+$$

and note that

$$Q_2 < R \Longleftrightarrow L_1(T_1) > \frac{1}{g}\left(s - \frac{R}{\tau_1}\right) \stackrel{def}{=} b_1$$

Denote by $\mathcal{V}$ the path value of the coupons that depend on the first knockout event

$$\mathcal{V} = \sum_{n=2}^{N-1} \frac{1}{B(T_{n+1})} X_n(T_n) \mathbb{I}_{\{Q_n < R\}}$$

and note that since $E[V|L_1(T_1) \leq b_1] = 0$ we have

$$E[\mathcal{V}] = E[\mathcal{V}|L_1(T_1) > b_1] Q^B(L_1(T_1) > b)$$

- Since T_1 is typically small, the probability $Q^B(L_1(T_1) > b)$ of not knocking out can be approximated analytically with high precision.
- $E[\mathcal{V}|L_1(T_1) > b_1]$ can be interpreted as the value of the TARN under the condition that it does not knock out on the date T_1, which is less discontinuous than our original payoff function.

More generally Piterbarg shows that all discontinuities can in fact be integrated outside of Monte Carlo.

Proposition 16.1. *Consider a TARN with structured coupon $C_n = (s - gL_n(T_n))^+$ and denote $b_n = \frac{1}{g}\left(s - \frac{R-Q_n}{\tau_n}\right)$. Its value is given by*

$$V_{tarn}(0) = \sum_{n=1}^{N-1} \tilde{E}^B \left[\psi_n E_{T_{n-1}} \left[\frac{X_n(T_n)}{B(T_{n+1})}\right]\right], \quad (148)$$

$$\psi_n = \prod_{k=1}^{n-1} Q^B_{T_{k-1}}(L_k(T_k) - b_k) \quad (149)$$

where the measure $\tilde{Q}^B$ is defined by the Radon-Nikodym derivative process $\Lambda(t) = E_t\left[\frac{d\tilde{Q}^B}{dQ^B}\right]$ where

$$\Lambda(t) = \frac{Q^B_t(L_{m+1}(T_{m+1}) > b_{m+1})}{Q^B_{T_m}(L_{m+1}(T_{m+1}) > b_{m+1})} \prod_{k=1}^{m-1} \frac{\mathbb{I}_{\{L_{m+1}(T_{m+1}) > b_{m+1}\}}}{Q^B_{T_k}(L_{m+1}(T_{k+1}) > b_{k+1})}, \quad t \in [T_m, t_{m+1}) \quad (150)$$

- The quantities ψ_n in (149) can be approximated analytically since each term of the form $Q^B_{T_{k-1}}(L_k(T_k > b_k)$ involves an expected value over a short period $[T_{k-1}, T_k]$, so short-time (e.g. Gaussian) approximations to the distribution of L_k over $[T_{k-1}, T_k]$ may be applied effectively (in fact, in some simulation schemes the ψ_n's come for free).
- Note that the functions ψ_n replace the indicator functions $\mathbb{I}_{\{Q_n < R\}}$ as weights on the coupons in the TARN valuation formula. The former are much smoother functions of a simulated path than are indicator functions.
- In measure $\tilde{Q}^B$, the TARN never knocks out. However, in order to use the method in practice we need to establish how to simulate model dynamics in this measure.

16.2.2.1 Simulating under survival measure using conditional Gaussian draws

Consider the case of a single-factor Libor market model

$$dL_i(t) = \lambda_i(t)\varphi(L_i(t))\left[\mu_i(t)dt + dW^B(t)\right], \quad i = 1, \ldots, N-1$$

In order to implement TARN pricing formula (148) we need to simulate Libor rates under the measure $\tilde{Q}^B$, namely, in such a way that $L_n(T_n) > b_n$ for each n. Consider a simulation step from T_{n-1} to T_n. Employing an Euler scheme we can approximate that Q^B-dynamics as

$$L_n(T_n) = L_n(T_{n-1}) + \lambda_n(T_{n-1})\varphi(L_n(T_{n-1}))\left[\mu_n(T_{n-1})\tau_{n-1} + \sqrt{\tau_{n-1}}Z\right], \quad Z \sim \mathcal{N}(0,1)$$

In order to ensure that $L_n(T_n) > b_n$ we need to make sure that Z satisfies

$$L_n(T_{n-1}) + \lambda_n(T_{n-1})\varphi(L_n(T_{n-1}))\left[\mu_n(T_{n-1})\tau_{n-1} + \sqrt{\tau_{n-1}}Z\right] > b_n$$

namely

$$Z > Z_{min} \stackrel{def}{=} \frac{b_n - m_n}{v_n}, \qquad \begin{cases} v_n = \lambda_n(T_{n-1})\varphi(L_n(T_{n-1}))\sqrt{\tau_{n-1}}, \\ m_n = L_n(T_{n-1}) + v_n\mu_n(T_{n-1})\sqrt{\tau_{n-1}} \end{cases}$$

so that $L_n(T_n) = m_n + v_n Z$. In order to simulate $\tilde{Z}$ conditioned on it being above a certain level Z_{min} we proceed as above

$$\tilde{Z} = \mathcal{N}^{-1}\left[\mathcal{N}(Z_{min}) + (1 - \mathcal{N}(Z_{min}))U\right], \quad U \sim \mathcal{U}(0,1)$$

It now suffices to verify that this measure just constructed actually coincides with the measure defined by the Radon-Nikodym process (150) (c.f. [AP10-3, Page 1071]).

- Once $\tilde{Z}$ has been drawn, all Libor rates can be evaluated using

$$L_i(T_n) = L_i(T_{n-1}) + \lambda_i(T_{n-1})\varphi(L_i(T_{n-1}))\left[\mu_i(T_{n-1})\tau_{n-1} + \sqrt{\tau_{n-1}}\tilde{Z}\right]$$

- The weights ψ_n in (149) come for free within this framework, since

$$Q^B_{T_{n-1}}(L_n(T_n) > b_n) = 1 - \mathcal{N}(Z_{min})$$

16.3 Control variates

Suppose we wish to estimate $E[Y]$ by Monte Carlo; the idea of the control variates method is to is a possibly multi-dimensional random variable with known mean which is closely related to Y in order to reduce the variance of the simulation. Concretely, if $\boldsymbol{Y^c} = (Y_1^c, \ldots, Y_q^c)^\top$ is a q-dimensional vector of random variables with known mean $\boldsymbol{\mu^c} = E[\boldsymbol{Y^c}] = (\mu_1^c, \ldots, \mu_q^c)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_q)^\top$ is a constant vector, we may form a new random variable

$$X = Y - \boldsymbol{\beta}^\top \cdot (\boldsymbol{Y^c} - \boldsymbol{\mu^c})$$

Clearly $E[X] = E[Y]$ so we can run a Monte Carlo simulation on X to estimate the mean of Y.

Denote the $q \times q$ covariance matrix of $\boldsymbol{Y^c}$ by

$$\boldsymbol{\Sigma}_{Y^c}, \qquad [\boldsymbol{\Sigma}_{Y^c}]_{i,j} = \mathrm{Cov}(Y_i^c, Y_j^c)$$

and define the q-dimensional vector

$$\boldsymbol{\Sigma}_{YY^c}, \qquad [\boldsymbol{\Sigma}_{YY^c}]_i = \mathrm{Cov}(Y, Y_i^c)$$

Then clearly

$$\mathrm{Var}(X) = \mathrm{Var}(Y) + \boldsymbol{\beta}^\top \Sigma_{Y^c}\boldsymbol{\beta} - 2\boldsymbol{\beta}^\top \Sigma_{YY^c}$$

Lemma 16.2. *$Var(X)$ is minimized at $\boldsymbol{\beta}^* = \Sigma_{Y^c}^{-1}\Sigma_{YY^c}$ and the minimum value is given by*

$$\min_{\boldsymbol{\beta}} Var(X) = (1 - R^2)\, Var(Y), \quad R^2 = \frac{1}{Var(Y)}\Sigma_{YY^c}^{\top}\Sigma_{Y^c}^{-1}\Sigma_{YY^c}$$

where we identify R^2 as the R-squared of a multi-dimensional regression of Y against Y^c and the components of the optimal vector $\boldsymbol{\beta}^*$ as the corresponding regression coefficients.

The ratio of variances is hence $\frac{\min_\beta \mathrm{Var}X}{\mathrm{Var}Y} = 1 - R^2$ so the larger the correlation between the Y and the control variate, the larger the variance reduction.

In our usual derivative pricing setting, we seek to replace our standard Monte Carlo estimate

$$\frac{1}{K}\sum_{j=1}^{K} Y(\omega_j)$$

with

$$\frac{1}{K}\sum_{j=1}^{K} \left(Y(\omega_j) - \boldsymbol{\beta}^{\top}(Y^c(\omega_j) - E[Y^c])\right)$$

where $Y^c(\omega_j)$ are random samples of the multi-dimensional control variate Y^c, chosen such that $E[Y^c]$ is available in closed form.

If Y is the value of a security under a given model, there are multiple ways to select the control variate Y^c.

- Y^c may represent the value of the same security under a different (but closely related) model.
- Y^c may represent the value of a different, but closely related, security under the same model.
- Y^c may represent the value of an approximate hedging strategy.
- Y^c may be a weighted combination of any of the above control variates.

We explore some of these in the next subsections, but we first outline a few elementary examples from [Gla03, Section 4.1].

Example 16.5 (pricing complex options using more tractable options as variates)**.** Consider pricing an arithmetic Asian option with payoff $\bar{S}_A = \frac{1}{n}\sum_{i=1}^{n} S(t_i)$ in the Black-Scholes model. No closed-form solution is known, so one must rely on simulation. In contrast, calls and puts on the geometric average $\bar{S}_G = \left(\prod_{i=1}^{n} S(t_i)\right)^{1/n}$ can be priced in closed form and are good candidates[45] to be used as control variates in pricing options on $\bar{S}_A$. For an arithmetic Asian call option, one could form a controlled estimator using independent replications of

$$(\bar{S}_A(\omega) - K)^+ - b\left\{(\bar{S}_G(\omega) - K)^+ - E[e^{-rT}(\bar{S}_G - K)^+]\right\}$$

Example 16.6 (pricing options using prices in simpler models as variates)**.** Consider pricing an option on an asset with dynamics

$$\frac{dS(t)}{S(t)} = rdt + \sigma(t)dW(t)$$

We can simulate S at dates $t_1, \ldots, t_n$ using

$$S(t_{i+1}) = S(t_i)\exp\left([r - \frac{1}{2}\sigma(t_i)^2](t_{i+1} - t_i) + \sigma(t_i)\sqrt{t_{i+1} - t_i}Z_{i+1}\right), \quad Z_i \sim \mathcal{N}(0,1)$$

[45] [Gla03, Figure 4.2] shows that there is an almost perfect correlation between the payoff of a call option on arithmetic average and the payoff of a call option on the geometric average.

Assume the option in question had a more tractable price if the asset were a geometric Brownian motion. We could simulate

$$\tilde{S}(t_{i+1}) = \tilde{S}(t_i)\exp\left([r - \frac{1}{2}\tilde{\sigma}^2](t_{i+1} - t_i) + \tilde{\sigma}\sqrt{t_{i+1} - t_i}Z_{i+1}\right), \quad Z_i \sim \mathcal{N}(0,1)$$

for some constant $\tilde{\sigma}$ and initial condition $\tilde{S}(0) = S(0)$. For instance, for a standard call stuck at K, we could form a controlled estimator using independent replications of

$$(S(t_n) - K)^+ - b\left\{(\tilde{S}(t_n)_K)^+ - E[(\tilde{S}(t_n) - K)^+]\right\}$$

For effective variance reduction, one should choose $\tilde{\sigma}$ to be close to a typical value of σ.

16.3.1 Model-based control variates

We fix some notation:

- Let $\hat{V}_{orig}$ be a Monte Carlo estimate for the true security value in the original pricing model
- Let $\hat{V}_{proxy}$ be a Monte Carlo estimate for the same security in a proxy model where a highly accurate price estimate $E[\hat{V}_{proxy}]$ is available.
- Denote $V_{PDE} = E[\hat{V}_{proxy}]$.

We then introduce a corrected value estimate as

$$\hat{V}_{corrected} = \hat{V}_{orig} - \beta\left(\hat{V}_{proxy} - V_{PDE}\right)$$

where β is the appropriate regression coefficient. If the path values used to compute $\hat{V}_{orig}$ are positively correlated with the ones used to obtain $\hat{V}_{proxy}$, the variance of the estimate is reduced.

We illustrate **how to construct a PDE-friendly model proxy for the LM model**. Recall that the LM model is Markovian only in the full set of all forward Libor rates on the yield curve, plus any additional variables required to model unspanned SV. For concreteness, consider a one-factor model with deterministic local volatility

$$\begin{aligned} dL_n(t) &= \varphi(L_n(t))\lambda_n(t)^\top\left(\mu_n(t,\boldsymbol{L}(t))dt + dW^B(t)\right) \\ \mu_n(t,\boldsymbol{L}(t)) &= \sum_{j=q(t)}^{n}\frac{\tau_j\varphi(L_j(t))\lambda_j(t)}{1+\tau_j L_j(t)} \end{aligned}$$

We seek to construct a low-dimensional Markovian approximation in a number of steps.

1. Eliminate the local volatility $\varphi(L_n(t))$ via the transform

$$f(x) = \int_{x_0}^{x}\frac{1}{\varphi(\xi)}\,d\xi, \quad \ell_n(t) \stackrel{def}{=} f(L_n(t)), \quad n = 1,\ldots,N-1$$

   which eliminates φ from the diffusion part of the SDE and yields

$$d\ell_n(t) = \lambda_n(t)^\top\left[\left(\mu_n(t,\boldsymbol{L}(t)) - \frac{1}{2}\lambda_n(t)\varphi'(L_n(t))\right)dt + dW(t)\right]$$

2. The LM model drift terms are generally small, so we may approximate

$$\mu_n(t,\boldsymbol{L}(t)) \simeq \mu_n(t,\boldsymbol{L}(0))$$

3. For local volatility functions $\varphi(x)$ which are close to linear, we may further approximate

$$\varphi'(L_n(t)) \simeq \varphi'(L_n(0))$$

which yields

$$d\ell_n(t) = \lambda_n(t)^\top \left[\left(\mu_n(t, \boldsymbol{L}(0)) - \frac{1}{2}\lambda_n(t)\varphi'(L_n(0)) \right) dt + dW(t) \right]$$

so that each $\ell_n(t)$ is an integral of $\lambda_n(t)$ against a Brownian motion with deterministic drift.

4. To make all variables functions of the same state variable, approximate the volatility structure with a separable one

$$\lambda_n(t) \simeq \hat{\lambda}_n(t) = \sigma_n \alpha(t), \quad n = 1, \ldots, N-1$$

where α is a function of time common to all Libor rates and σ_n's are Libor-specific scalars. Defining a one-dimensional Markovian state variable by

$$dX(t) = \alpha(t) dW(t)$$

we obtain that all the variables $\ell_n(t)$ are deterministic functions of $X(t)$:

$$\begin{aligned} \ell_n(t) &= \ell_n(0) + d_n(t) + \sigma_n X(t), \\ d_n(t) &= \int_0^t \lambda_n(s) \cdot \left(\mu_n(s, \boldsymbol{L}(0)) - \frac{1}{2}\lambda_n(s)\varphi'(L_n(0)) \right) ds \end{aligned}$$

which translated back into Libor forwards reads

$$L_n(t) = f^{-1}\left(f(L_n(0)) + d_n(t) + \sigma_n X(t) \right), \quad n = 1, \ldots, N-1$$

5. With this representation, at each point in time t, the value of any path-independent derivative V can be expressed as a function

$$V = V(t, X(t))$$

satisfying the PDE

$$\frac{\partial V(t,x)}{\partial t} + \frac{1}{2}\alpha^2(t)\frac{\partial^2 V(t,x)}{\partial x^2} = r(t,x)V(t,x)$$

subject to appropriate boundary and jump conditions.

To use the Markov approximation as the control variate, we define:

- $\hat{V}_{proxy}$: value of the security in the Markov LM model computed by Monte Carlo.
- V_{PDE}: value of the security computed by the PDE method in the same model.

In order to achieve the required high correlation between the path values of a derivative in the original and proxy models, it is necessary to use the same simulation seed for random number generation and also compatible discretization schemes.

The *potential downside of the method* is the fact that its scope is limited by the need to perform a PDE valuation, with three dimensions being the practical maximum for a reasonably quick PDE scheme.

16.3.2 Instrument-based control variates

The idea here is to keep the model fixed but change the payoff; we illustrate how the method works when applied to Bermudan swaptions and CLE's.

Consider a Bermudan swaption with $N-1$ exercise opportunities $T_1, \ldots, T_{N-1}$, with corresponding exercise values

$$U_n(t) = B(t)E_t\left[\sum_{i=n}^{N-1} \frac{1}{B(T_{i+1})} X_i\right], \quad X_i = \tau_i(L_i(T_i) - \kappa)$$

The value of the Bermudan swaption is given by

$$H_0(0) = E\left[\sum_{i=\eta}^{N-1} \frac{1}{B(T_{i+1})} X_i\right]$$

where η is the optimal exercise index, and the corresponding K-path Monte Carlo estimate value is

$$H_0(0) \simeq \frac{1}{K}\sum_{k=1}^{K} Y(\omega_k), \qquad Y(\omega) = \sum_{i=\eta(\omega)}^{N-1} \frac{1}{B(T_{i+1}, \omega)} X_i(\omega)$$

where $\omega_1, \ldots, \omega_K$ are simulated paths.

A naive set of control variates that we could introduce is based on the exercise values

$$Y_n^c(\omega) = U_n(T_n, \omega) = B(T_n)\left[\sum_{i=n}^{N-1} \frac{1}{B(T_{i+1}, \omega)} X_i(\omega)\right]$$

but this presents two major drawbacks:

1. The control variates Y_n^c, $n = 1, \ldots, N-1$ are linear functions of rates, whereas the payoff of a Bermudan swaption is not well approximated by a linear function.
2. More subtly, the value of a Bermudan swaption (or CLE) along a path ω involves cash-flows fixed at times $T_{\eta(\omega)}, \ldots, T_N$, whereas all the cash-flows in the definition of Y_n^c are sampled at a single time. Since the number of coupons included in the Bermudan swaption and the controls differ, the correlation between them is likely to be low.

16.3.2.1 Timing mismatch

The timing mismatch in (2) can be rectified by sampling the controls at the Bermudan swaption exercise time, since:

Lemma 16.3. *Let $U_n = U_n(T_n)$ and let Z_n, $n = 0, \ldots, N-1$ be a collection of martingales with respect to the filtration $\{\mathcal{F}_{T_n}\}_{n=0}^N$. If $\eta, \sigma \in \{1, \ldots, N-1\}$ are stopping times satisfying $\eta \le \sigma$, then*

$$[Corr(U_\eta, Z_\eta)]^2 \ge [Corr(U_\eta, Z_\sigma)]^2$$

Therefore if a European option maturing at T_n is used as a control variate, and for a particular Monte Carlo path $\eta < n$, the Lemma implies that we should use the value of the option at time T_η to generate a control variate (higher correlation) rather than wait until the maturity date T_n. Namely, we could consider control variates

$$Y_n^c \stackrel{def}{=} U_n(T_\eta) = B(T_\eta)\sum_{i=n}^{\eta-1} \frac{X_i}{B(T_{i+1})} + B(T_\eta)E_{T_\eta}\left[\sum_{\max\{\eta,n\}}^{N-1} \frac{X_i}{B(T_{i+1})}\right], \quad n = 1, \ldots, N-1$$

Note also that by the optional sampling theorem $Y_n^c = U_n(T_\eta)$ are martingales, whence

$$E[Y_n^c] = E[U_n(T_\eta)] = U_n(0)$$

16.3.2.2 Non-linear controls

To introduce non-linearity into the controls for Bermudan swaptions, one can consider using caps, which can be valued exactly in the majority of LM models. Concretely, for a Bermuda swaption with strike κ whose exercise values are

$$U_n(t) = B(t) \sum_{i=n}^{N-1} E_t \left[\frac{1}{B(T_{i+1})} \tau_i (L_i(T_i) - \kappa) \right]$$

consider the set of caps

$$V_{cap,n}(t) = B(t) \sum_{i=n}^{N-1} E_t \left[\frac{1}{B(T_{i+1})} \tau_i (L_i(T_i) - \kappa)^+ \right], \quad n = 1, \ldots, N-1$$

These can be used to construct a few possible controls. For instance, we can use a collection of all caps, observed at the Bermudan exercise time

$$\begin{aligned} \boldsymbol{Y}^c &= (Y_1^c, \ldots, Y_{N-1}^c)^\top, \\ Y_n^c &= V_{cap,n}(T_\eta) = B(T_\eta) \sum_{i=n}^{\eta-1} \frac{1}{B(T_{i+1})} \tau_i (L_i(T_i) - \kappa)^+ \\ &\quad + B(T_\eta) E_{T_\eta} \left[\sum_{\max\{\eta,n\}}^{N-1} \frac{1}{B(T_{i+1})} \tau_i (L_i(T_i) - \kappa)^+ \right] \end{aligned}$$

For securities more complex than the standard Bermudan swaptions, the idea of sampling the controls at the exercise time still applies, but the choice of controls becomes non-trivial and has to be done on a case-by-case basis.

16.3.3 Dynamic control variates

We present here the *delta* method of Clewlow and Carverhill. Let $X(t)$ be a p-dimensional vector process al of whose components are martingales (e.g. assets deflated by a numraire) and $g : \mathbb{R}^p \to \mathbb{R}$ a smooth function. Consider estimating the expected value

$$V(0) = E[g(X(T))]$$

Under regularity conditions we have

$$V(T) = V(0) + \int_0^T \sum_{i=1}^{p} \frac{\partial V(t)}{\partial X_i(t)} \, dX_i(t)$$

On a simulation time line $\{t_j\}_{j=1}^m$ this can be written in the style of an Euler scheme

$$V(T) = V(0) + \sum_{j=1}^{m} \sum_{i=1}^{p} \frac{\partial V(t_{j-1})}{\partial X_i(t)} \left(X_i(t_j) - X_i(t_{j-1}) \right).$$

The zero-mean quantity

$$\sum_{j=1}^{m} \sum_{i=1}^{p} \frac{\partial V(t_{j-1})}{\partial X_i(t)} \left(X_i(t_j) - X_i(t_{j-1}) \right)$$

is likely to have high correlation to $V(T)$, so we can consider using it as a control variate. In practice, one can use deltas constructed using approximate risk sensitivities in order build the control variate.

- For the LM model, deltas could originate from an approximate Markov model as the one in Section 16.3.1.
- In [AP10-3, Section 25.5], a general technique due to Moni is described which is suitable for CLE's, which we outline below.

Take as regression variables polynomials of explanatory variables $\boldsymbol{x}(t) = (x_1(t), \ldots, x_d(t))$ and approximate the hold value using least squares

$$\tilde{H}_n(T_n) = p_n(\boldsymbol{x}(T_n)), \quad n = 0, \ldots, N-1$$

Approximate sensitivities can then be computed as

$$\frac{\Delta H_n(T_n)}{\Delta x_m(T_n)} \simeq \frac{\Delta \tilde{H}_n(T_n)}{\Delta x_m(T_n)} = \frac{\partial p_n}{\partial x_m}(\boldsymbol{X}(T_n)), \quad m = 1, \ldots, d$$

which yields an approximate hedging strategy

$$V_{hs}(T_n) = H_0(0) + \sum_{j=0}^{n-1} \left(\sum_{m=0}^{d} \frac{\partial p_n}{\partial x_m}(\boldsymbol{X}(T_j)) \left[x_m(T_{j+1}) - x_m(T_j)\right] \right), \quad n = 1, \ldots, N-1$$

with time-0 expected value

$$E\left[V_{hs}(T_n)\right] = H_0(0) + \sum_{j=0}^{n-1} E\left[\sum_{m=0}^{d} \frac{\partial p_n}{\partial x_m}(\boldsymbol{X}(T_j)) \left(E[x_m(T_{j+1})|\mathcal{F}_{T_j}] - x_m(T_j)\right) \right] \overset{[*]}{=} H_0(0)$$

where in $[*]$ we assumed that we selected the explanatory variable to be martingales, whereby $E[x_m(T_{j+1})|\mathcal{F}_{T_j}] = x_m(T_j)$.

A control variate can then be defined as the hedging strategy stopped at the exercise time

$$Y^c \overset{def}{=} V_{hs}(T_\eta)$$

which according to tests by Moni yields reductions in standard error by a factor of 2 to 3.

References

[AP10-1] L. Andersen, V. Piterbarg, *Interest rate modeling.* Volume 1: Foundations and vanilla models, Atlantic Financial Press, 2010

[AP10-2] L. Andersen, V. Piterbarg, *Interest rate modeling.* Volume 2: Term structure models, Atlantic Financial Press, 2010

[AP10-3] L. Andersen, V. Piterbarg, *Interest rate modeling.* Volume 3: Products and risk management, Atlantic Financial Press, 2010

[Gla03] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, 2003

[Shr04] S. E. Shreve, *Stochastic calculus for finance II*, Springer, 2004