Volatility Surface of Foreign Exchange

Copyright © Changwei Xiong 2016 - 2024

November 14, 2016

last update: June 6, 2024

TABLE OF CONTENTS

Table of Contents 2
1. Mathematical Preliminaries 5
1.1. Kolmogorov Forward and Backward Equations 5
1.1.1. Kolmogorov Forward Equation 5
1.1.2. Kolmogorov Backward Equation 7
1.2. Tanaka's Formula 9
1.3. Generalized Gyöngy Theorem 10
1.4. Approximation by Discrete Markov Chain 11
1.4.1. Ornstein-Uhlenbeck Process 16
1.4.2. Wiener Process 18
2. FX Option Market Conventions 19
2.1. Option Trading Strategies 19
2.1.1. Single Call and Put 19
2.1.2. Call Spread and Put Spread 19
2.1.3. Risk Reversal, Straddle and Strangle 19
2.1.4. Butterfly 20
2.2. Black-Scholes Formula 20
2.3. Foreign-Domestic Symmetry 23
2.4. Market Quoting Convention 23
2.5. Risk Sensitivities 24
2.5.1. Delta 24
2.5.1.1. Pips Spot Delta 24
2.5.1.2. Percentage Spot Delta 25
2.5.1.3. Pips Forward Delta 27
2.5.1.4. Percentage Forward Delta 27
2.5.1.5. Strike from Delta Conversion 27
2.5.2. Other Risk Sensitivities 29
2.5.2.1. Theta 29
2.5.2.2. Gamma 29
2.5.2.3. Vega 30
2.5.2.4. Vanna 30
2.5.2.5. Volga 31
2.6. FX Volatility Smile Convention 31
2.6.1. At-The-Money Volatility 31
2.6.1.1. ATM Forward 32

2.6.1.2. Delta Neutral Straddle 32
2.6.2. Risk Reversal Volatility 33
2.6.3. Strangle Volatility 33
2.6.3.1. Market Strangle 33
2.6.3.2. Smile Strangle 34
2.6.4. Smile Volatility 35
3. Volatility Surface Construction 36
3.1. Smile Interpolation 36
3.1.1. Polynomial-in-Delta 36
3.1.2. Stochastic Volatility Inspired (SVI) 38
3.1.2.1. Uni-SVI 39
3.1.2.2. Bi-SVI 40
3.1.2.3. Tri-SVI 41
3.2. Temporal Interpolation 41
3.3. Volatility Surface by Standard Conventions 41
4. The Vanna-Volga Method 43
4.1. Vanna-Volga Pricing 43
4.2. Smile Interpolation 44
5. Classic Local Volatility: Dupire 48
5.1. Local Volatility by Vanilla Call 48
5.2. Local Volatility by Undiscounted Call 51
5.3. Local Volatility by Implied Volatility 52
5.3.1. Formula in Log Spot Moneyness 54
5.3.2. Formula in Log Forward Moneyness 54
5.3.3. Conversion between Log Forward Moneyness and Log Spot Moneyness 56
5.3.4. Equivalency in Formulas 56
5.4. Forward Smile in Local Volatility 58
6. Local Volatility with Stochastic Rates: General Dupire 61
6.1. General Dupire Local Volatility 62
6.2. The General Dupire Model 66
6.3. Stochastic Rates: One-Factor Hull White Model 67
6.3.1. Model Value of Swaption: Jamshidian Decomposition 69
6.3.2. Market Value of Swaption 72
6.4. Transition Probability Density Function 72
6.5. Model Calibration by Forward PDE 74
6.5.1. Numerical Solution of Forward PDE 75

6.6. Pricing by Backward PDE....77
7. Stochastic Local Volatility: Ornstein-Uhlenbeck Dupire....80
8. Stochastic Local Volatility: Markov Chain Ornstein-Uhlenbeck Dupire....80
9. Stochastic Local Volatility: Mixed Dupire....80
10. Stochastic Local Volatility with Stochastic Rates: Mixed General Dupire....80
11. Local Volatility: PDE by Finite Difference Method....81
11.1. Date Conventions of Equity and Equity Option....81
11.2. Deterministic Dividends....82
11.3. Forward PDE....83
11.3.1. Treatment of Deterministic Dividends....84
11.4. Backward PDE....86
11.4.1. PDE in Centered Log-spot....86
11.4.1.1. Treatment of Deterministic Dividends....87
11.4.1.2. Vanilla Call....88
11.4.2. PDE in Log-spot....88
11.4.2.1. Treatment of Deterministic Dividends....89
11.5. Local Volatility Surface....89
11.6. Barrier Option Pricing....90
References....92

1. Mathematical Preliminaries

1.1. Kolmogorov Forward and Backward Equations

The time evolution of the transition probability density function is governed by *Kolmogorov* forward and backward equations, which will be introduced as follows, without loss of generality, in multi-dimension.

1.1.1. Kolmogorov Forward Equation

Let's consider the following m-dimensional stochastic spot process $X_t \in \mathbb{R}^m$ driven by an n-dimensional Brownian motion W_t whose correlation matrix ρ is given by $\rho dt = dW_t dW_t'$

$$\underset{m\times 1}{dX_t} = \underset{m\times 1}{A(t,X_t)}\underset{1\times 1}{dt} + \underset{m\times n}{B(t,X_t)}\underset{n\times 1}{dW_t} \tag{1}$$

We derive the dynamics of h, where $h: \mathbb{R}^m \longrightarrow \mathbb{R}$ is a scalar-valued Borel-measurable function only on variable X_t

$$\underset{1\times 1}{dh(X_t)} = \underset{1\times m}{J_h}\underset{m\times 1}{dX_t} + \frac{1}{2}\underset{1\times m}{dX_t'}\underset{m\times m}{H_h}\underset{m\times 1}{dX_t} = J_h A dt + J_h B dW_t + \frac{1}{2} dW_t' B' H_h B dW_t \tag{2}$$

where J_h is the $1 \times m$ *Jacobian* (i.e., the same as *gradient* if h is a scalar-valued function) and H_h the $m \times m$ *Hessian* (with subscripts now denoting the indices of vector components)

$$J_h = \begin{pmatrix} \dfrac{\partial h}{\partial X_1} & \cdots & \dfrac{\partial h}{\partial X_m} \end{pmatrix}, \qquad H_h = \begin{pmatrix} \dfrac{\partial^2 h}{\partial X_1^2} & \cdots & \dfrac{\partial^2 h}{\partial X_1 \partial X_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 h}{\partial X_m \partial X_1} & \cdots & \dfrac{\partial^2 h}{\partial X_m^2} \end{pmatrix} \tag{3}$$

Expanding the expression in (2), we have

$$\begin{aligned} dh(X_t) &= \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} A_i dt + \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} \sum_{k=1}^{n} B_{ik} dW_{k,t} + \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2 h}{\partial X_i \partial X_j} \sum_{k=1}^{n} B_{ik}\rho_{ij}B_{jk} dt \\ &= \left(\sum_{i=1}^{m} A_i \frac{\partial h}{\partial X_i} + \frac{1}{2} \sum_{i,j=1}^{m} \Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right) dt + \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} \sum_{k=1}^{n} B_{ik} dW_{k,t} \end{aligned} \tag{4}$$

where $\Sigma = B\rho B'$ is the $m \times m$ instantaneous variance-covariance matrix of dX. Integrating on both sides of (4) from initial time s to time t, we have

$$h(X_t) - h(X_s) = \int_s^t \left(\sum_{i=1}^{m} A_i \frac{\partial h}{\partial X_i} + \frac{1}{2} \sum_{i,j=1}^{m} \Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right) du + \int_s^t \sum_{i=1}^{m} \frac{\partial h}{\partial X_i} \sum_{k=1}^{n} B_{ik} dW_{k,t} \tag{5}$$

Taking expectation on both sides of (5), we get (using notation $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_t]$)

$$\begin{aligned} \text{LHS} &= \mathbb{E}_s[h(X_t)] - h(X_s) = \int_\Omega h_x p_{t,x|s,\alpha} dx - h_\alpha \\ \text{RHS} &= \int_s^t \sum_{i=1}^{m} \mathbb{E}_s\left[A_i \frac{\partial h}{\partial X_i} \right] du + \frac{1}{2} \int_s^t \sum_{i,j=1}^{m} \mathbb{E}_s\left[\Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right] du \end{aligned} \tag{6}$$

where $p_{t,x|s,\alpha}$ is the transition probability density function having $X_t = x$ at t given $X_s = \alpha$ at s (i.e., if we solve the equation (1) with the initial condition $X_s = \alpha \in \mathbb{R}^m$, then the random variable $X_t = x \in \Omega$ has a density $p_{t,x|s,\alpha}$ in the x variable at time t). Differentiating (6) with respect to t on both sides, we get

$$\begin{aligned} \int_\Omega h_x \frac{\partial p_{t,x|s,\alpha}}{\partial t} dx &= \sum_{i=1}^{m} \mathbb{E}_s\left[A_i \frac{\partial h}{\partial X_i} \right] + \frac{1}{2} \sum_{i,j=1}^{m} \mathbb{E}_s\left[\Sigma_{ij} \frac{\partial^2 h}{\partial X_i \partial X_j} \right] \\ &= \sum_{i=1}^{m} \int_\Omega p_{t,x|s,\alpha} A_i \frac{\partial h_x}{\partial x_i} dx + \frac{1}{2} \sum_{i,j=1}^{m} \int_\Omega p_{t,x|s,\alpha} \Sigma_{ij} \frac{\partial^2 h_x}{\partial x_i \partial x_j} dx \end{aligned} \tag{7}$$

If we assume $\Omega \equiv \mathbb{R}^m$ and also assume the probability density p and its first derivatives $\partial p / \partial x_i$ vanish at a higher order of rate than h and $\partial h / \partial x_i$ as $x_i \to \pm\infty \; \forall \; i = 1, \cdots, m$, we can integrate by parts for the right hand side of (7), once for the first integral and twice for the second

$$\int_\Omega A_i p \frac{\partial h_x}{\partial x_i} dx = \int_{\bar{\Omega}_i} \underbrace{A_i h_x p|_{x_i=-\infty}^{+\infty}}_{=0} d\bar{x}_i - \int_\Omega h_x \frac{\partial (A_i p)}{\partial x_i} dx \tag{8}$$

$$\int_\Omega \Sigma_{ij}\frac{\partial^2 h_x}{\partial x_i \partial x_j} p dx = \int_{\bar{\Omega}_i} \underbrace{\Sigma_{ij}\frac{\partial h_x}{\partial x_j} p\bigg|_{x_i=-\infty}^{+\infty}}_{=0} d\bar{x}_i - \int_\Omega \frac{\partial(\Sigma_{ij}p)}{\partial x_i}\frac{\partial h_x}{\partial x_j} dx$$

$$= -\int_{\bar{\Omega}_j} \underbrace{h_x \frac{\partial(\Sigma_{ij}p)}{\partial x_i}\bigg|_{x_j=-\infty}^{+\infty}}_{=0} d\bar{x}_j + \int_\Omega h_x \frac{\partial^2(\Sigma_{ij}p)}{\partial x_i \partial x_j} dx$$

where $$\int_{\bar{\Omega}_i} (\cdot) d\bar{x}_i = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} (\cdot) dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_m$$

Plugging the results of (8) into (7), we have

$$\begin{aligned} &\int_\Omega h_x \frac{\partial p}{\partial t} dx = -\sum_{i=1}^m \int_\Omega h_x \frac{\partial(A_i p)}{\partial x_i} dx + \frac{1}{2}\sum_{i,j=1}^m \int_\Omega h_x \frac{\partial^2(\Sigma_{ij}p)}{\partial x_i \partial x_j} dx \\ &\Rightarrow \int_\Omega h_x \left(\frac{\partial p}{\partial t} + \sum_{i=1}^m \frac{\partial(A_i p)}{\partial x_i} - \frac{1}{2}\sum_{i,j=1}^m \frac{\partial^2(\Sigma_{ij}p)}{\partial x_i \partial x_j}\right) dx = 0 \end{aligned} \tag{9}$$

By the arbitrariness of function h, we conclude that for any $x \in \Omega$ the density function p satisfies

$$\frac{\partial p}{\partial t} + \sum_{i=1}^m \frac{\partial(A_i p)}{\partial x_i} - \frac{1}{2}\sum_{i,j=1}^m \frac{\partial^2(\Sigma_{ij}p)}{\partial x_i \partial x_j} = 0, \qquad \Sigma = B\rho B' \tag{10}$$

This is the multi-dimensional *Fokker-Planck Equation (a.k.a. Kolmogorov Forward Equation)* [1]. In this equation, the s and α are held constant, while the t and x are variables (called “forward variables”). In the one-dimensional case, it reduces to

$$\frac{\partial p}{\partial t} + \frac{\partial(Ap)}{\partial x} - \frac{1}{2}\frac{\partial^2(B^2 p)}{\partial x^2} = 0 \tag{11}$$

where $A = A(t,x)$ and $B = B(t,x)$ are then scalar functions.

1.1.2. Kolmogorov Backward Equation

Let’s express conditional expectation $g(t,X_t) = \mathbb{E}_t[h(X_T)]$. Since for any $t \le v \le T$ we have

$$g(t,X_t) = \mathbb{E}_t[h(X_T)] = \mathbb{E}_t\big[\mathbb{E}_v[h(X_T)]\big] = \mathbb{E}_t[g(v,X_v)] \tag{12}$$

the $g(t, X_t)$ is a martingale by the *tower rule* (i.e., If $\mathcal{H}$ holds less information than $\mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$). The dynamics of the $g(t, X_t)$ is given by

$$dg = \frac{\partial g}{\partial t}dt + \underset{1\times m}{J_g}\underset{m\times 1}{dX_t} + \frac{1}{2}\underset{1\times m}{dX_t'}\underset{m\times m}{H_g}\underset{m\times 1}{dX_t} = \frac{\partial g}{\partial t}dt + J_g A dt + J_g B dW_t + \frac{1}{2}dW_t' B' H_g B dW_t \quad (13)$$

where J_g is the *Jacobian* and H_g the *Hessian* of g with respect to variable X

$$[J_g]_i = \frac{\partial g}{\partial X_i}, \qquad [H_g]_{ij} = \frac{\partial^2 g}{\partial X_i \partial X_j} \quad (14)$$

Expanding (13), we have

$$dg = \left(\frac{\partial g}{\partial t} + \sum_{i=1}^{m} A_i \frac{\partial g}{\partial X_i} + \frac{1}{2}\sum_{i,j=1}^{m} \Sigma_{ij}\frac{\partial^2 g}{\partial X_i \partial X_j}\right)dt + \sum_{i=1}^{m}\frac{\partial g}{\partial X_i}\sum_{k=1}^{n} B_{ik} dW_{k,t} \quad (15)$$

Since $g(t, X_t)$ is a martingale, the dt-term must vanish, which gives

$$\frac{\partial g}{\partial t} + \sum_{i=1}^{m} A_i \frac{\partial g}{\partial X_i} + \frac{1}{2}\sum_{i,j=1}^{m} \Sigma_{ij}\frac{\partial^2 g}{\partial X_i \partial X_j} = 0 \quad (16)$$

This is the multi-dimensional *Feynman-Kac* formula[1].

Using the transition probability density $p_{T,\beta|t,x}$ for $X_t = x$ at t and $X_T = \beta$ at T, we can further write the expectation as

$$g_{t,x} = \mathbb{E}_t[h(X_T)] = \int_{\Omega} h_\beta p_{T,\beta|t,x} d\beta \quad (17)$$

The formula (16) defines that

$$\begin{aligned} &\left(\frac{\partial}{\partial t} + \sum_{i=1}^{m} A_i \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m} \Sigma_{ij}\frac{\partial^2}{\partial x_i \partial x_j}\right)\int_{\Omega} h_\beta p_{T,\beta|t,x} d\beta = 0 \\ &\Rightarrow \int_{\Omega} h_\beta \left(\frac{\partial p}{\partial t} + \sum_{i=1}^{m} A_i \frac{\partial p}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{m} \Sigma_{ij}\frac{\partial^2 p}{\partial x_i \partial x_j}\right) d\beta = 0 \end{aligned} \quad (18)$$

By the arbitrariness of h function, we must have

[1] https://en.wikipedia.org/wiki/Feynman-Kac_formula

$$\frac{\partial p}{\partial t}+\sum_{i=1}^{m}A_i\frac{\partial p}{\partial x_i}+\frac{1}{2}\sum_{i,j=1}^{m}\Sigma_{ij}\frac{\partial^2 p}{\partial x_i\partial x_j}=0,\qquad \Sigma=B\rho B' \tag{19}$$

This is the multi-dimensional *Kolmogorov Backward Equation*. In this equation, the T and β are held constant, while the t and x are variables (called "backward variables"). In the 1-D case, it reduces to

$$\frac{\partial p}{\partial t}+A\frac{\partial p}{\partial x}+\frac{1}{2}B^2\frac{\partial^2 p}{\partial x^2}=0 \tag{20}$$

where $A=A(t,x)$ and $B=B(t,x)$ are again scalar functions.

1.2. Tanaka's Formula

Tanaka's formula can be seen as the analogue of Itō's lemma for the (nonsmooth) absolute value function $f(X)=|X|$ with $\frac{\partial f}{\partial X}=\mathrm{Sgn}(X)$ and $\frac{\partial^2 f}{\partial X^2}=2\delta(X)$. Assume that X_t is a semimartingale (e.g., Ito processes, which satisfy a stochastic differential equation of the form $dX_t=a_t dt+b_t dW_t$, are semimartingales), then for every fixed $K\in\mathbb{R}$

$$\begin{aligned}
&|X_t-K|=|X_s-K|+\int_s^t \mathrm{Sgn}(X_u-K)dX_u+L_t^K(X),\qquad \mathrm{Sgn}(x)=\begin{cases}+1, & x>0\\ 0, & x=0\\ -1, & x<0\end{cases}\\
&L_t^K(X)=\int_s^t\delta(X_u-K)d\langle X,X\rangle_u=\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_s^t \mathbb{1}_{\{X_u\in(K-\varepsilon,K+\varepsilon)\}}d\langle X,X\rangle_u
\end{aligned} \tag{21}$$

where $d\langle X,X\rangle_u$ is the quadratic variation and $L_t^K(X)$ is the local time spent by X_t around K between s and t, which can be thought of *occupation density* of process X_t around K during that period [2] [3]. For a standard Brownian motion B_t, its local time reads

$$\begin{aligned}
&|B_t-K|=|B_s-K|+\int_s^t\mathrm{Sgn}(B_u-K)dB_u+L_t^K(B)\\
&L_t^K(B)=\int_s^t\delta(B_u-K)du=\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_s^t\mathbb{1}_{\{B_u\in(K-\varepsilon,K+\varepsilon)\}}du\\
&\qquad\qquad=\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\left|\{u\in[s,t]|B_u\in(K-\varepsilon,K+\varepsilon)\}\right|
\end{aligned} \tag{22}$$

In differential form, (21) transforms into

$$d|X_t - K| = \mathrm{Sgn}(X_t - K)dX_t + dL_t^K(X), \qquad dL_t^K(X) = \delta(X_t - K)d\langle X, X\rangle_t \tag{23}$$

where the differential operates on the t-variable. Further, notice that the terminal payoff function of call or put option can be written as

$$(X_t - K)^+ = \frac{|X_t - K| + X_t - K}{2}, \qquad (K - X_t)^+ = \frac{|X_t - K| + K - X_t}{2} \tag{24}$$

We may derive the dynamics of the payoff function as

$$\begin{aligned} d(X_t - K)^+ &= \frac{1}{2}d|X_t - K| + \frac{1}{2}dX_t = \frac{1}{2}\mathrm{Sgn}(X_t - K)dX_t + \frac{1}{2}\delta(X_t - K)d\langle X, X\rangle_t + \frac{1}{2}dX_t \\ &= \frac{2\Theta(X_t - K) - 1}{2}dX_t + \frac{1}{2}\delta(X_t - K)d\langle X, X\rangle_t + \frac{1}{2}dX_t \\ &= \Theta(X_t - K)dX_t + \frac{1}{2}\delta(X_t - K)d\langle X, X\rangle_t \\ d(K - X_t)^+ &= \frac{1}{2}d|X_t - K| - \frac{1}{2}dX_t = \frac{1}{2}\mathrm{Sgn}(X_t - K)dX_t + \frac{1}{2}\delta(X_t - K)d\langle X, X\rangle_t - \frac{1}{2}dX_t \\ &= (\Theta(X_t - K) - 1)dX_t + \frac{1}{2}\delta(X_t - K)d\langle X, X\rangle_t \\ &= -\Theta(K - X_t)dX_t + \frac{1}{2}\delta(K - X_t)d\langle X, X\rangle_t \end{aligned} \tag{25}$$

where Θ is the *Heaviside* step function[1] and δ is the *Dirac* delta function[2].

1.3. Generalized Gyöngy Theorem

Let W be an N-dimensional Brownian motion, and

$$dX(t) = \mu(t)dt + \sigma(t)dW(t) \tag{26}$$

[1] Heaviside step function: $\Theta(x) = \begin{cases} 0, & x < 0 \\ 1/2 & x = 0 \\ 1, & x > 0 \end{cases}$. It can be viewed informally as the integral of the Dirac delta function: $\Theta(x) = \int_{-\infty}^{x} \delta(u)du$.

[2] Dirac delta function: $\delta(x) = \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$ and subject to constraint $\int_{-\infty}^{\infty} \delta(u)du = 1$. It can be viewed informally as the derivative of the Heaviside step function: $\delta(x) = d\Theta(x)/dx$.

be a K-dimensional Ito process where $\mu(t)$ is a bounded K-dimensional adapted process, and $\sigma(t)$ is a bounded $K \times N$-dimensional adapted process such that $\sigma(t)\sigma(t)'$ is uniformly positive definite. There exist deterministic measurable functions $\hat{\mu}$ and $\hat{\sigma}$ such that

$$\begin{aligned} &\hat{\mu}(t,x) = \mathbb{E}[\mu(t)|X(t) = x] \\ &\hat{\sigma}(t,x)\hat{\sigma}(t,x)' = \mathbb{E}[\sigma(t)\sigma(t)'|X(t) = x] \end{aligned} \tag{27}$$

and there exists a weak solution to the stochastic differential equation

$$d\hat{X}(t) = \hat{\mu}\big(t,\hat{X}(t)\big)dt + \hat{\sigma}\big(t,\hat{X}(t)\big)d\widehat{W}(t) \tag{28}$$

such that the Markov process $\hat{X}(t)$ admits the same marginal probability distribution as that of $X(t)$ for every $t > 0$. The $\widehat{W}$ denotes another Brownian motion, possibly on another space. This is the generalized Gyöngy theorem in multi-dimension [4] [5] [6] [7] [8].

Put it simply, Gyöngy theorem states that a given stochastic differential equation (SDE) with stochastic drift and diffusion coefficients, can be mimicked by another constructed process, which has deterministic coefficients. The solutions of the two equations have the same marginal probability distributions. It links local volatility models in the form of (28) to other diffusion models in the more general form of (26) that are capable of generating the same implied volatility surface. The local volatility model is in some sense the simplest among the diffusion models capable of reproducing the implied volatility surface. This is in analogy to the fact that there can be countless 3D objects with different shapes casting however identical shadows on the ground. As in (27), local volatility $\hat{\sigma}(t,x)$ can be regarded as the conditional risk-neutral expectation of the instantaneous variance of the asset $X(t)$, given that the asset level at the future time t is x. This is analogous to the known relationship between the forward and future spot interest rates where the forward rate is the forward-risk-adjusted expectation of the instantaneous future spot rate. The local volatility represents some kind of average over all possible instantaneous volatilities at a certain point in time in a stochastic volatility world [9].

1.4. Approximation by Discrete Markov Chain

To simplify a model, we sometimes seek to approximate a continuous (Markovian) stochastic process by a Markov chain process on a discrete set of states which jumps with certain jump intensities between neighboring states. Suppose there are N states placed a uniform width a_t apart and centered about zero, we have a discrete stochastic process Λ_t such that

$$\Lambda_t = i_t a_t \quad \text{for} \quad i_t \in \{-m, -m+1, \cdots, m-1, m\}, \qquad m = \frac{N-1}{2} \tag{29}$$

where i_t denotes the discrete state recorded at time t. The Λ_t process jumps between neighboring points with rightwards (increasing value) intensities $(N-1, N-2, \cdots, 1, 0)h_t$ and with leftwards (decreasing value) intensities $(0, 1, \cdots, N-2, N-1)h_t$ where h_t is a unit intensity to be calibrated. In other words, if we are in the leftmost state, e.g., $i_t = -m$, we may only jump one step right, and do so with intensity $(N-1)h_t$. If we are in the next state along, we may move one step left with intensity h_t or one step right with intensity $(N-2)h_t$, and so on. The key property of these intensities is that the total jump intensity in each state is $(N-1)h_t$ with only the probability of going in each direction changing. Below we summarize the probability $\mathbb{P}(i_{t+\Delta t}|i_t)$ of a jump from state i_t to $i_{t+\Delta t}$ in an infinitesimal increment of time Δt

$$i_t \longrightarrow i_{t+\Delta t} = \begin{cases} i_t - 1, & \mathbb{P}(i_{t+\Delta t}|i_t) = (m + i_t)h_t \Delta t \\ i_t, & \mathbb{P}(i_{t+\Delta t}|i_t) = 1 - 2m h_t \Delta t \\ i_t + 1, & \mathbb{P}(i_{t+\Delta t}|i_t) = (m - i_t)h_t \Delta t \end{cases} \tag{30}$$

After sufficient time, the process can reach an equilibrium at which the occupation probability of each state becomes stationary (i.e., time-invariant). Let x be the stationary probability vector. By definition, we have

$$x(k) = \mathbb{P}_t(k) = \lim_{t-s\to\infty} \mathbb{P}[i_t = k | i_s = p], \qquad p, k \in \{-m, \cdots, m\} \tag{31}$$

The vector-valued stationary probability $\mathbb{P}_t(k)$ is characterized by the Markov chain transition matrix M of the process, which is derived from (30). Its (i,j)-entry defines the transition probability from state j to state i within time Δt

$$M(i,j)=\begin{cases}(m+j)h_t\Delta t, & \text{If } i=j-1 \quad \text{(super diagonal)}\\ 1-2mh_t\Delta t, & \text{If } i=j \quad \text{(main diagonal)}\\ (m-j)h_t\Delta t, & \text{If } i=j+1 \quad \text{(sub diagonal)}\\ 0, & \text{Otherwise}\end{cases},\qquad i,j\in\{-m,\cdots,m\} \tag{32}$$

Mathematically speaking, the stationary probability vector x is an eigenvector of the transition matrix M associated with the eigenvalue 1, e.g., $Mx=x$. For a matrix with strictly positive entries (or, more generally, for an irreducible aperiodic stochastic matrix), the stationary probability vector x is unique and can be computed by observing that for any j we have the following limit

$$\lim_{k\to\infty} M^k(i,j)=x(i) \tag{33}$$

In our model, the state occupation at equilibrium follows a binomial distribution $B(N-1,1/2)$ with probability mass function

$$\mathbb{P}_t(k)=\frac{1}{2^{2m}}\cdot\frac{(2m)!}{(m+k)!\,(m-k)!},\qquad k\in\{-m,-m+1,\cdots,m-1,m\} \tag{34}$$

This can be proved by finding an equilibrium in the inflow and outflow of the k-th state occupation probability at time t for an infinitesimal time interval Δt, that is

$$\begin{aligned}\mathbb{P}^{\text{Inflow}}_{t\to t+\Delta t}(k)&=M(k,k-1)\mathbb{P}_t(k-1)+M(k,k)\mathbb{P}_t(k)+M(k,k+1)\mathbb{P}_t(k+1)\\ \mathbb{P}^{\text{Outflow}}_{t\to t+\Delta t}(k)&=M(k-1,k)\mathbb{P}_t(k)+M(k,k)\mathbb{P}_t(k)+M(k+1,k)\mathbb{P}_t(k)\end{aligned} \tag{35}$$

where

$$\begin{aligned}&M(k,k-1)=(m-k+1)h_t\Delta t,\qquad M(k,k+1)=(m+k+1)h_t\Delta t\\ &M(k-1,k)=(m+k)h_t\Delta t,\qquad M(k+1,k)=(m-k)h_t\Delta t\end{aligned} \tag{36}$$

The probability net flow at the k-th state within time Δt can then be calculated as

$$\begin{aligned}&\mathbb{P}^{\text{Inflow}}_{t\to t+\Delta t}(k)-\mathbb{P}^{\text{Outflow}}_{t\to t+\Delta t}(k)\\ &=M(k,k-1)\mathbb{P}_t(k-1)+M(k,k+1)\mathbb{P}_t(k+1)-M(k-1,k)\mathbb{P}_t(k)-M(k+1,k)\mathbb{P}_t(k)\end{aligned} \tag{37}$$

$$= \frac{(2m)!}{2^{2m}} h_t \Delta t \left(\frac{m-k+1}{(m+k-1)!\,(m-k+1)!} + \frac{m+k+1}{(m+k+1)!\,(m-k-1)!} - \frac{m+k}{(m+k)!\,(m-k)!} - \frac{m-k}{(m+k)!\,(m-k)!} \right)$$

$$= 0$$

Since the probability net flow is zero for any of the states (i.e., $Mx = x$) at time t, this shows that state occupation at equilibrium follows a binomial distribution with probability mass function (34).

Knowing the probability mass function (34), we can compute the mean and variance of the Λ_t process, providing the binomial theorem, which says

$$(\alpha+\beta)^p = \sum_{k=0}^{p} \frac{p!}{k!\,(p-k)!} \alpha^k \beta^{p-k} \tag{38}$$

By defining $m = n/2$ and $i = j - n/2$, we can derive its mean as

$$\begin{aligned} \mathbb{E}[\Lambda_t] &= \sum_{i=-m}^{m} i a_t \mathbb{P}_t(i) = a_t \sum_{i=-m}^{m} \frac{i}{2^{2m}} \frac{(2m)!}{(m+i)!\,(m-i)!} = a_t \sum_{j=0}^{n} \frac{j-\frac{n}{2}}{2^n} \frac{n!}{j!\,(n-j)!} \\ &= a_t \sum_{j=0}^{n} \frac{j}{2^n} \frac{n!}{j!\,(n-j)!} - \frac{n a_t}{2} \sum_{j=0}^{n} \frac{1}{2^n} \frac{n!}{j!\,(n-j)!} = a_t \sum_{j=1}^{n} \frac{1}{2^n} \frac{n!}{(j-1)!\,(n-j)!} - \frac{n a_t}{2} \\ &= a_t \frac{n}{2} \sum_{k=0}^{p} \frac{1}{2^p} \frac{p!}{k!\,(p-k)!} - \frac{n a_t}{2} = 0, \qquad \text{with } p = n-1 \text{ and } k = j-1 \end{aligned} \tag{39}$$

and its variance as

$$\begin{aligned} \mathbb{V}[\Lambda_t] &= \mathbb{E}[\Lambda_t^2] - (\mathbb{E}[\Lambda_t])^2 = \mathbb{E}[\Lambda_t^2] = \mathbb{E}[\Lambda_t(\Lambda_t - 1)] = \sum_{i=-m}^{m} i(i-1) a_t^2 \mathbb{P}_t(i) \\ &= a_t^2 \sum_{i=-m}^{m} \frac{i(i-1)}{2^{2m}} \frac{(2m)!}{(m+i)!\,(m-i)!} = a_t^2 \sum_{j=0}^{n} \frac{\left(j-\frac{n}{2}\right)\left(j-\frac{n}{2}-1\right)}{2^n} \frac{n!}{j!\,(n-j)!} \\ &= a_t^2 \sum_{j=0}^{n} \frac{j^2 - nj + \frac{n^2}{4} - j + \frac{n}{2}}{2^n} \frac{n!}{j!\,(n-j)!} = a_t^2 \sum_{j=2}^{n} \frac{1}{2^n} \frac{n!}{(j-2)!\,(n-j)!} - \frac{n^2 a_t^2}{2} + \frac{n^2 a_t^2}{4} + \frac{n a_t^2}{2} \end{aligned} \tag{40}$$

$$= \frac{n(n-1)a_t^2}{4} \sum_{k=0}^{p} \frac{1}{2^p} \frac{p!}{k!\,(p-l)!} - \frac{n^2 a_t^2}{4} + \frac{n a_t^2}{2} = \frac{m a_t^2}{2}, \qquad \text{with } p = n-2 \text{ and } k = j-2$$

We can also work out the expectation and variance of infinitesimal change in discrete stochastic process Λ_t. By noting that the change in Λ_t may come from both the change in a_t and the jump in i_t, we have

$$\begin{aligned} \mathbb{E}_t[d\Lambda_t] &= i_t da_t + a_t \mathbb{E}_t[di_t] = i_t da_t - 2h_t \Lambda_t dt = \left(\frac{1}{a_t}\frac{da_t}{dt} - 2h_t\right)\Lambda_t dt \\ \mathbb{V}_t[d\Lambda_t] &= \mathbb{E}_t[(d\Lambda_t - \mathbb{E}_t[d\Lambda_t])^2] = \mathbb{E}_t[(a_t di_t + 2h_t \Lambda_t dt)^2] = a_t^2 \mathbb{E}_t[(di_t)^2] = 2m h_t a_t^2 dt \end{aligned} \tag{41}$$

where we further derive

$$\begin{aligned} \mathbb{E}_t[di_t] &= \mathbb{E}_t[i_{t+dt} - i_t] \\ &= (i_t - 1)(m + i_t)h_t dt + i_t(1 - 2m h_t dt) + (i_t + 1)(m - i_t)h_t dt - i_t \\ &= -2 i_t h_t dt \\ \mathbb{E}_t[(di_t)^2] &= \mathbb{E}_t[(i_{t+dt} - i_t)^2] = \mathbb{E}_t[i_{t+dt}^2] - 2 i_t \mathbb{E}_t[i_{t+dt}] + i_t^2 \\ &= (i_t - 1)^2(m + i_t)h_t dt + i_t^2(1 - 2m h_t dt) + (i_t + 1)^2(m - i_t)h_t dt \\ &\quad - 2 i_t(i_t - 2 i_t h_t dt) + i_t^2 = 2m h_t dt \end{aligned} \tag{42}$$

By matching $\mathbb{E}_t[d\Lambda_t]$ and $\mathbb{V}_t[d\Lambda_t]$ to their counterparts, we can approximate a continuous stochastic process by the discrete process Λ_t.

A good analogy for the jumps in the discrete process would be the failures in reliability theory. We may assume that the probability of a failure that will occur in an infinitesimal time interval $[t, t + \Delta t]$ is $h_t \Delta t$ with intensity (i.e., hazard rate) h_t, that is

$$h_t \Delta t = \mathbb{P}[\tau < t + \Delta t | \tau > t] = \frac{\mathbb{P}[t < \tau < t + \Delta t]}{\mathbb{P}[\tau > t]} = \frac{\mathbb{P}[\tau < t + \Delta t] - \mathbb{P}[\tau < t]}{\mathbb{P}[\tau > t]} \tag{43}$$

where τ denote the time when a failure occurs and $\mathbb{P}[\tau > t]$ is the survival probability. As $\Delta t \to 0$, we can write (43) as

$$h_t = -\lim_{\Delta t \to 0} \frac{\mathbb{P}[\tau > t + \Delta t] - \mathbb{P}[\tau > t]}{\mathbb{P}[\tau > t]\Delta t} = -\frac{d \log \mathbb{P}[\tau > t]}{dt} \tag{44}$$

And the conditional survival probability would be

$$\mathbb{P}[\tau > t | \tau > s] = \frac{\mathbb{P}[\tau > t]}{\mathbb{P}[\tau > s]} = \exp\left(-\int_s^t h_u du\right) \tag{45}$$

which will be used later to calculate jump probability between two jump times.

1.4.1. Ornstein-Uhlenbeck Process

For a simple *Ornstein-Uhlenbeck* Process Y_t below, we can find its solution together with its mean and variance as follows

$$\begin{aligned}
&dY_t = -\kappa Y_t dt + \zeta dZ_t, \qquad Y_t = e^{-\kappa(t-v)} Y_v + \zeta \int_v^t e^{-\kappa(t-u)} dZ_u, \qquad Y_s = \mathcal{N}(0, \gamma^2) \\
&\mathbb{E}_v[Y_t] = e^{-\kappa(t-v)} Y_v, \qquad \mathbb{E}[Y_t] = 0, \qquad \mathbb{V}[Y_t] = e^{-2\kappa(t-s)}\gamma^2 + \frac{\zeta^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right) \equiv \nu_t \\
&\lim_{\kappa \to 0} \mathbb{V}[Y_t] = \gamma^2 + \zeta^2(t-s), \qquad \lim_{t \to \infty} \mathbb{V}[Y_t] = \frac{\zeta^2}{2\kappa} \\
&\mathbb{E}_t[dY_t] = -\kappa Y_t dt, \qquad \mathbb{V}_t[dY_t] = \zeta^2 dt, \qquad s < v < t
\end{aligned} \tag{46}$$

We want to match the discrete Λ_t process to the continuous *Ornstein-Uhlenbeck* process Y_t in (46) by equating the mean and variance in (41) to the counterparts in (46), that is, we want to have $\mathbb{E}_t[d\Lambda_t] = \mathbb{E}_t[dY_t]$ and $\mathbb{V}_t[d\Lambda_t] = \mathbb{V}_t[dY_t]$. This gives a first order linear ordinary differential equation below, whose solution can be obtained given that the discrete process must reproduce the variance of the continuous process, i.e., $\mathbb{V}[\Lambda_t] = \mathbb{V}[Y_t] = \nu_t$ with $\mathbb{V}[\Lambda_t]$ in (40)

$$\begin{aligned}
&\mathbb{E}_t[d\Lambda_t] = \left(\frac{1}{a_t}\frac{da_t}{dt} - 2h_t\right)\Lambda_t dt = -\kappa \Lambda_t dt, \qquad \mathbb{V}_t[d\Lambda_t] = 2mh_t a_t^2 dt = \zeta^2 dt \\
&\Longrightarrow \frac{1}{a_t}\frac{da_t}{dt} = 2h_t - \kappa, \qquad 2mh_t a_t^2 = \zeta^2 \Longrightarrow \frac{da_t^2}{dt} = -2\kappa a_t^2 + \frac{2\zeta^2}{m} \\
&\Longrightarrow a_t^2 = \frac{2\nu_t}{m}, \qquad h_t = \frac{\zeta^2}{4\nu_t}
\end{aligned} \tag{47}$$

However, the jump intensity h_t in (47) will blow up when t approaches zero if the initial uncertainty γ^2 is zero, rendering the transition matrix ill-posed. We deal with this by assuming zero transition intensities until some time has passed, which is numerically acceptable as prices should not be particularly sensitive to the volatility dynamics over the first week or so. Since h_t is a function of t, the transition probability for a small timestep from v to t for $s < v < t$ can be integrated analytically as below

$$\int_v^t h_u du = \frac{\zeta^2}{4}\int_v^t \frac{1}{\nu_u} du = \frac{1}{4}\log\left(\frac{\gamma^2 + \frac{\zeta^2}{2\kappa}\left(e^{2\kappa(t-s)} - 1\right)}{\gamma^2 + \frac{\zeta^2}{2\kappa}\left(e^{2\kappa(v-s)} - 1\right)}\right) = \frac{\kappa(t-v)}{2} + \frac{1}{4}\log\frac{\nu_t}{\nu_v} \tag{48}$$

There are two ways to handle the discrete jump process numerically: 1) evolve the process with transition probability over a sequence of time intervals (with error $\sim\Delta t$), and 2) directly model the jump times, t_j. For PDE implementation, it is necessary to use the first method. For Monte Carlo simulation, the second is more efficient. Noting that the total jump intensity is $2mh_t$, which is state independent, the jump probability from last jump time t_j to next jump time t_{j+1} can be derived from (45) as

$$\mathbb{P}\left[\tau < t_{j+1} \middle| \tau > t_j\right] = 1 - \exp\left(-2m\int_{t_j}^{t_{j+1}} h_u du\right) \tag{49}$$

We can simulate the next jump time by drawing a uniform random number $z \sim U(0,1)$. The next jump time t_{j+1}, which is random, can be derived from (48) and (49) by

$$\mathbb{P}\left[\tau < t_{j+1} \middle| \tau > t_j\right] = z \Longrightarrow t_{j+1} = s + \frac{1}{2\kappa}\log\left(\frac{e^{2\kappa(t_j - s)} + \frac{2\kappa\gamma^2}{\zeta^2} - 1}{(1-z)^{\frac{2}{m}}} - \frac{2\kappa\gamma^2}{\zeta^2} + 1\right) \tag{50}$$

Once the next jump time is simulated, one can draw another uniform random number to find if the state should jump left or right according to the respective jump intensities.

1.4.2. Wiener Process

To match the discrete Λ_t process to a Wiener process W_t (i.e., standard Brownian motion process), we would require that

$$\begin{aligned}
&\mathbb{E}_t[d\Lambda_t] = \left(\frac{d\log a_t}{dt} - 2h_t\right)\Lambda_t dt = 0, \qquad \mathbb{V}_t[d\Lambda_t] = 2mh_t a_t^2 dt = dt \\
&\Rightarrow \; \frac{d\log a_t}{dt} = 2h_t, \qquad 2mh_t a_t^2 = 1 \;\Rightarrow\; a_t^2 = \frac{2(t-s)}{m}, \qquad h_t = \frac{1}{4(t-s)}
\end{aligned} \tag{51}$$

From (40), we see that $\mathbb{V}[\Lambda_t] = \mathbb{V}[W_t] = t - s$.

In addition, using (49) and the integral

$$\int_t^T h_u du = \frac{1}{4}\int_t^T \frac{1}{u-s} du = \frac{1}{4}\log\left(\frac{T-s}{t-s}\right) \tag{52}$$

we can derive the next jump time t_{j+1} for simulation as

$$t_{j+1} = s + \frac{t_j - s}{(1-z)^{\frac{2}{m}}} \tag{53}$$

where $z \sim U(0,1)$ is again a uniform random number.

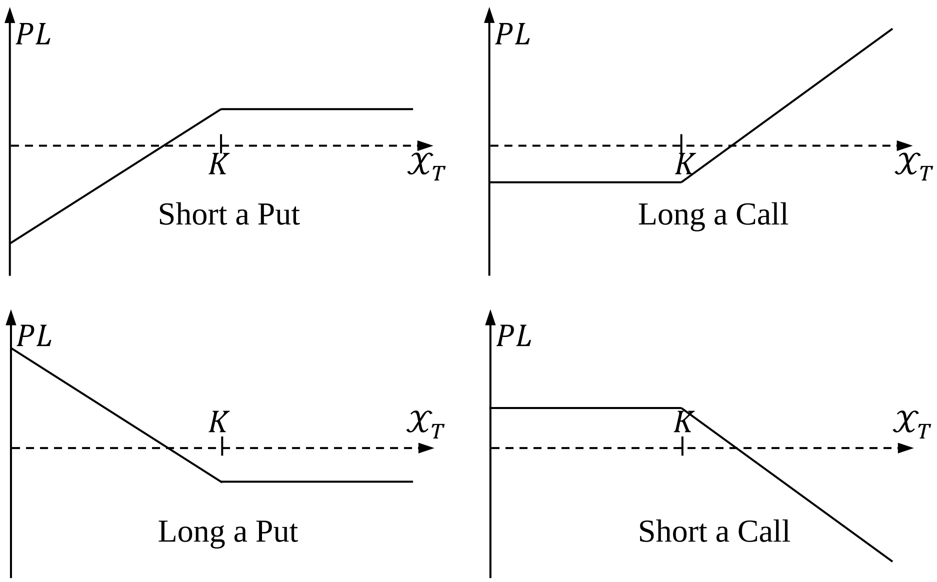
2. FX Option Market Conventions

2.1. Option Trading Strategies

In the following, we will introduce a few simple trading strategies of vanilla options, which are are liquidly traded in FX markets. Letting $\mathcal{X}_t$ denote the FX spot observed at time t and K the strike value, we may plot the payoff of each instrument as a function of the terminal FX spot $\mathcal{X}_T$ level as follows.

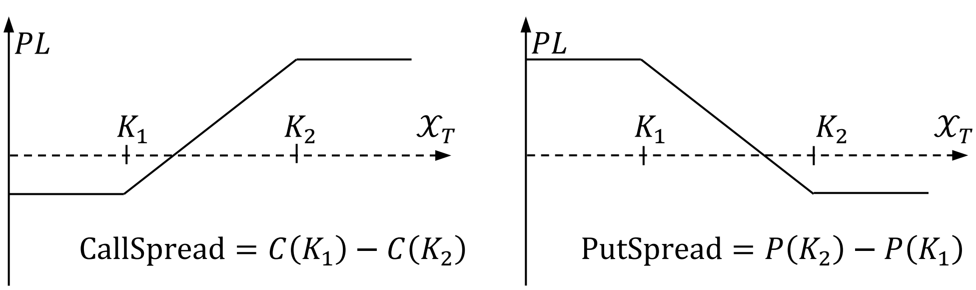
2.1.1. <u>Single Call and Put</u>

The figures below depict the payoff functions of vanilla options.



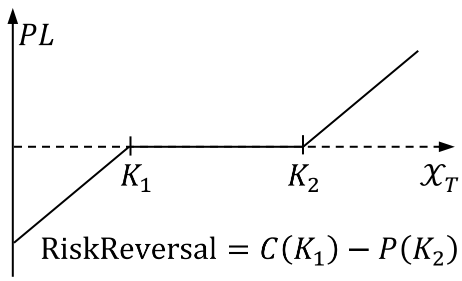
2.1.2. <u>Call Spread and Put Spread</u>

A call spread is a combination of a long call and a short call option with different strikes $K_1 < K_2$. A put spread is a combination of a long put and a short put option with different strikes. The figure below shows the payoff functions of a call spread and a put spread.

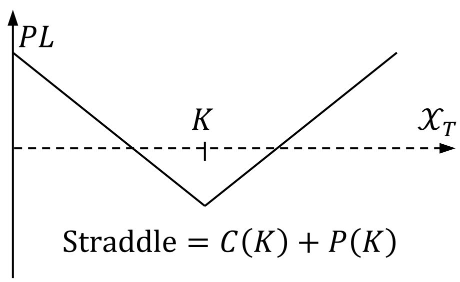


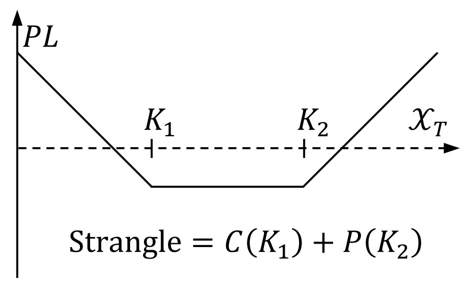
2.1.3. <u>Risk Reversal, Straddle and Strangle</u>

A risk reversal (RR) is a combination of a long call and a short put with different strikes $K_1 < K_2$. This is a zero-cost product as one can finance a call option by short selling a put option. The figure below shows the payoff function of a risk reversal.



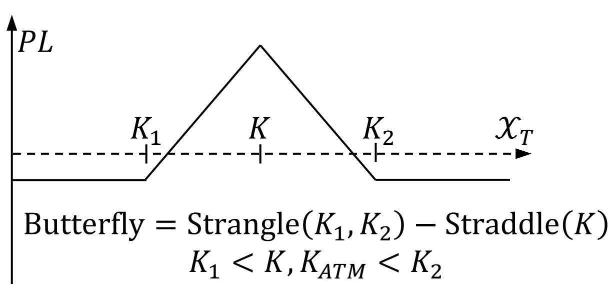
A straddle is a combination of a call and a put option with the same strike K. A strangle is a combination of an out-of-money call and an out-of-money put option with two different strikes $K_1 < K_{ATM} < K_2$. The figure below shows the payoff functions of a straddle and a strangle





2.1.4. Butterfly

A butterfly (BF) is combinations of a long strangle and a short straddle. The figure below shows the payoff function of a butterfly



2.2. Black-Scholes Formula

Currency pairs are commonly quoted using ISO codes in the format FORDOM, where FOR and DOM denote foreign and domestic currency respectively. For example, in EURUSD, the EUR denotes the foreign currency or currency1 and USD the domestic currency or currency2. The rate of EURUSD tells the price of one euro in USD.

In Black-Scholes model, FX spot rate is assumed to follow a geometric Brownian motion. Under domestic risk neutral measure, the FX spot is characterized by the following stochastic differential equation with a drift $r - \hat{r}$ and a volatility σ

$$\frac{d\mathcal{X}_t}{\mathcal{X}_t} = (r - \hat{r})dt + \sigma d\widetilde{W}_t \tag{54}$$

where the r and $\hat{r}$ are the domestic and foreign risk free rate respectively (the accent hat "^" here denotes a counterpart in the foreign economy, e.g., $\hat{r}$ is the foreign risk free rate). With the assumption of deterministic interest rates, an option on the FX spot with a strike K can be valued in Black model as

$$\begin{aligned} V &= P\mathfrak{B}(\omega, K, \sigma, \tau) = \omega\hat{P}\mathcal{X}\Phi(\omega d_+) - \omega PK\Phi(\omega d_-) \\ V_F &= \mathfrak{B}(\omega, K, \sigma, \tau) = \omega F\Phi(\omega d_+) - \omega K\Phi(\omega d_-) \end{aligned} \tag{55}$$

where V is the present value, V_F undiscounted value, $\omega = 1$ or -1 for call or put, $\tau = T - t$ for term to maturity, Φ for standard normal cumulative density function, and d_+ and d_- as follows

$$d_- = \frac{1}{\sigma\sqrt{\tau}}\log\frac{F}{K} - \frac{\sigma\sqrt{\tau}}{2} \qquad \text{and} \qquad d_+ = d_- + \sigma\sqrt{\tau} \tag{56}$$

In (55), the $P_{t,T}$ and $\hat{P}_{t,T}$ denote the domestic and the foreign zero coupon bond price (or equivalently the discount factors if rates are deterministic), respectively. The FX forward F is given by the *covered interest rate parity* (i.e., the returns from investing domestically must be equal to the returns from investing abroad to be arbitrage-free)

$$F_{t,T} = \mathcal{X}_t \frac{\hat{P}_{t,T}}{P_{t,T}}, \qquad P_{t,T} = \exp\left(-\int_t^T r_u du\right), \qquad \hat{P}_{t,T} = \exp\left(-\int_t^T \hat{r}_u du\right) \tag{57}$$

FX options are usually physically settled (i.e., upon exercise at maturity, the buyer of a EURUSD call receives notional N amount in EUR and pays NK amount in USD).

Black-Scholes pricing formula can be easily derived from arbitrage-free pricing

$$
\begin{aligned}
V_t &= M_t\widetilde{\mathbb{E}}_t\left[\frac{(\mathcal{X}_T-K)^+}{M_T}\right] = M_t\widetilde{\mathbb{E}}_t\left[\frac{\mathcal{X}_T\mathbb{1}\{\mathcal{X}_T>K\}}{M_T}\right] - KM_t\widetilde{\mathbb{E}}_t\left[\frac{\mathbb{1}\{\mathcal{X}_T>K\}}{M_T}\right] \\
&= X_t\mathbb{E}_t^X\left[\frac{\mathcal{X}_T\mathbb{1}\{\mathcal{X}_T>K\}}{X_t}\right] - KP_{t,T}\mathbb{E}_t^T\left[\frac{\mathbb{1}\{\mathcal{X}_T>K\}}{P_{T,T}}\right], \qquad \text{Change } M_t\to X_t=\widehat{M}_t\mathcal{X}_t,\ \ M_t\to P_{t,T} \\
&= \widehat{M}_t\mathcal{X}_t\mathbb{E}_t^X\left[\frac{\mathbb{1}\{\mathcal{X}_T>K\}}{\widehat{M}_T}\right] - KP_{t,T}\mathbb{E}_t^T[\mathbb{1}\{\mathcal{X}_T>K\}] \\
&= \widehat{P}_{t,T}\mathcal{X}_t\mathbb{E}_t^X[\mathbb{1}\{\mathcal{X}_T>K\}] - KP_{t,T}\mathbb{E}_t^T[\mathbb{1}\{\mathcal{X}_T>K\}], \qquad \text{assuming deterministic foreign rates} \\
&= \widehat{P}_{t,T}\mathcal{X}_t\mathbb{P}_t^X[\mathcal{X}_T>K] - KP_{t,T}\mathbb{P}_t^T[\mathcal{X}_T>K] \\
&= \widehat{P}_{t,T}\mathcal{X}_t\mathbb{P}_t^X[\mathcal{X}_T>K] - KP_{t,T}\widetilde{\mathbb{P}}_t[\mathcal{X}_T>K], \qquad \text{assuming deterministic domestic rates} \\
&= \widehat{P}\mathcal{X}\Phi(d_+) - PK\Phi(d_-)
\end{aligned}
\tag{58}
$$

where $\mathbb{P}_t^X[\mathcal{X}_T>K]$ and $\mathbb{P}_t^T[\mathcal{X}_T>K]$ are both conditional probabilities of spot finishing in-the-money at maturity. The $\mathbb{P}_t^X[\mathcal{X}_T>K]$ is computed under the measure associated with the foreign money market account denominated in domestic currency $X_t=\widehat{M}_t\mathcal{X}_t$ as the numeraire, whereas the $\mathbb{P}_t^T[\mathcal{X}_T>K]$ is computed under T-forward measure associated with domestic zero coupon bond $P_{t,T}$ as the numeraire. By assuming deterministic domestic interest rate, it will be equivalent to $\widetilde{\mathbb{P}}_t[\mathcal{X}_T>K]$ under risk neutral measure. Since the drift adjustment due to change of numeraire is

$$
\underset{\text{Under }\mathbb{Q}}{d\widetilde{W}_t} = \underset{\text{Under }\mathbb{N}}{dW_t^{\mathbb{N}}} + \sigma_N dt \tag{59}
$$

where $\mathbb{N}$ denotes the measure associated with numeraire N and $\mathbb{Q}$ the risk neutral measure. The FX spot process under the measure associated with the foreign money market account (basically, it is itself times the non-random $\widehat{M}_t$) as the numeraire is given by

$$
\frac{d\mathcal{X}_t}{\mathcal{X}_t} = (r-\hat{r})dt + \sigma d\widetilde{W}_t = (r-\hat{r}+\sigma^2)dt + \sigma dW_t^X \tag{60}
$$

The total drift adjustment $\sigma^2\tau$ for period $\tau = T - t$ is then normalized by the total volatility $\sigma\sqrt{\tau}$ of the exchange rate process $\mathcal{X}_t$ to give a shift term $\sigma\sqrt{\tau}$ as the difference between d_+ and d_- in the classic Black-Scholes formula.

2.3. Foreign-Domestic Symmetry

On top of the well-known put-call parity in options, there exists a foreign-domestic symmetry in currency options, shown as below

$$\frac{1}{\mathcal{X}} \cdot \mathrm{OptionValue}(\omega, \mathcal{X}, K, \sigma, r, \hat{r}, \tau) = K \cdot \mathrm{OptionValue}\left(-\omega, \frac{1}{\mathcal{X}}, \frac{1}{K}, \sigma, \hat{r}, r, \tau\right) \tag{61}$$

For example, a call on $\mathcal{X}$ is equivalent to a put on $\hat{\mathcal{X}} \equiv 1/\mathcal{X}$. Alternatively speaking, a right to buy one FOR at a price of K DOM is equivalent to the right to sell K DOM at a price of one FOR. In Black-Scholes model, the symmetry can be derived as follows, e.g., the value of a call on $\mathcal{X}$ is

$$V = P_{t,T}\big(F\Phi(d_+) - K\Phi(d_-)\big), \qquad d_+ = \frac{1}{\sigma\sqrt{\tau}}\log\frac{F}{K} + \frac{\sigma\sqrt{\tau}}{2}, \qquad d_- = d_+ - \sigma\sqrt{\tau} \tag{62}$$

and the value of a put on $\hat{\mathcal{X}}$ is

$$\begin{aligned} \hat{V} &= -\hat{P}_{t,T}\left(\hat{F}\Phi\big(-\hat{d}_+\big) - \hat{K}\Phi\big(-\hat{d}_-\big)\right) = \hat{P}_{t,T}\frac{F\Phi(d_+) - K\Phi(d_-)}{FK} = \frac{V}{\mathcal{X}K} \\ \hat{d}_+ &= \frac{1}{\sigma\sqrt{\tau}}\log\frac{\hat{F}}{\hat{K}} + \frac{\sigma\sqrt{\tau}}{2} = -d_-, \qquad \hat{d}_- = \hat{d}_+ - \sigma\sqrt{\tau} = -d_+ \end{aligned} \tag{63}$$

2.4. Market Quoting Convention

The option price quoting convention varies for currencies [10] [11]. Options can be quoted in one of the four relative quote styles: domestic per foreign (fd), percentage foreign ($\%f$), percentage domestic ($\%d$) and foreign per domestic (df). The call and put prices we showed in defined in (55) are actually expressed in domestic per foreign style (also known as the domestic pips price), denoted by V_{fd}. With the notional amount N expressed in foreign currency, we have $V_{fd} = NP\mathfrak{B}(\omega, K, \sigma, \tau)$. The other price quote styles have the following relationships with respect to V_{fd}

$$V_{\%f} = \frac{V_{fd}}{\mathcal{X}}, \qquad V_{\%d} = \frac{V_{fd}}{K}, \qquad V_{df} = \frac{V_{fd}}{\mathcal{X}K} \tag{64}$$

It is very important to note that this technique of constructing all these different quote styles only works where there are two notionals given by strike $K = N/\widehat{N}$, in foreign and domestic currencies, and there is a fixed relationship between them, which is known from the start. This is true for European and American style vanilla options, even in the presence of barriers and accrual features, but is most definitely not true for digital options. Suppose one has a cash-or-nothing digital which pays one USD if the EURUSD FX rate fixes at time T above a particular level (sometimes called 'strike', which actually leads to the confusion). The digital clearly has a USD notional (= $1, the domestic notional) so we can obtain percentage domestic (%USD) and foreign per domestic (EUR/USD) prices. However, there is no EUR notional (the foreign notional) at all so the other two quote styles are meaningless [12].

2.5. Risk Sensitivities

Risk sensitivity of an option is the sensitivity of the price to a change in underlying state variables or model parameters. We will present some basic types of risk sensitivities in the context of Black-Scholes model.

2.5.1. Delta

Delta is the ratio of change in option value to the change in spot or forward. There are several definitions of delta, such as spot/forward delta, pips/percentage delta, etc. Since FX volatility smiles are commonly quoted as a function of delta rather than as a function of strike, it is important to use a delta definition consistent with the market convention for the currency.

2.5.1.1. *Pips Spot Delta*

The pips spot delta is defined in Black-Scholes model as the first derivative of the present value with respect to the spot, both in domestic per foreign terms, corresponding to risk exposures in FOR. This style of delta implies that the premium currency is DOM and notional currency is FOR. It is commonly adopted by currency pairs with USD as DOM (or currency2), e.g., EURUSD, GBPUSD and AUDUSD, etc. By assuming $N=1$ in FOR and hence $V_{fd}=P\mathfrak{B}(\omega,K,\sigma,\tau)$, the pips spot delta is equivalent to the standard Black-Scholes delta

$$\Delta=\frac{\partial V_{fd}}{\partial \mathcal{X}}=\omega\hat{P}\Phi(\omega d_+)+\omega\hat{P}\mathcal{X}\phi(d_+)\omega\frac{\partial d_+}{\partial \mathcal{X}}-\omega PK\phi(d_-)\omega\frac{\partial d_-}{\partial \mathcal{X}}=\omega\hat{P}\Phi(\omega d_+) \tag{65}$$

where we have used the following identities

$$\begin{aligned}&\frac{\partial\Phi(\omega d_+)}{\partial d_+}=\omega\phi(d_+)=\frac{\omega}{\sqrt{2\pi}}\exp\left(-\frac{d_+^2}{2}\right),\qquad \frac{\partial\Phi(\omega d_-)}{\partial d_-}=\omega\phi(d_-)=\omega\frac{F}{K}\phi(d_+)\\&\Phi(-x)=1-\Phi(x),\qquad \frac{\partial d_+}{\partial \mathcal{X}}=\frac{\partial d_-}{\partial \mathcal{X}}=\frac{1}{\mathcal{X}\sigma\sqrt{\tau}}\end{aligned} \tag{66}$$

with ϕ the standard normal probability density function. To understand pips spot delta, assuming DOM is the numeraire, if one wants to hedge a short call of N notional in FOR with a premium of NV_{fd} in DOM, one must be long $N\Delta$ amount of the spot $\mathcal{X}$. This can be achieved by entering a long position of $N\Delta$ units of FOR with a cost of $N\Delta\mathcal{X}$ units of DOM.

2.5.1.2. *Percentage Spot Delta*

The percentage spot delta (also known as *premium adjusted* pips spot delta) is defined as a derivative of the present value with respect to the spot, both in percentage foreign terms, corresponding to risk exposures in DOM. This style of delta implies that the premium currency and notional currency both are FOR. It is used by currency pairs like USDJPY, EURGBP, etc. In Black-Scholes model, the percentage spot delta has the form

$$\Delta_{\%}=\frac{\partial V_{\%f}}{\frac{\partial \mathcal{X}}{\mathcal{X}}}=\mathcal{X}\frac{\partial}{\partial \mathcal{X}}\left(\frac{V_{fd}}{\mathcal{X}}\right)=\frac{\partial V_{fd}}{\partial \mathcal{X}}-\frac{V_{fd}}{\mathcal{X}}=\Delta_{\mathcal{X}}-V_{\%f}=\omega\hat{P}\frac{K}{F}\Phi(\omega d_-) \tag{67}$$

which shows that the percentage spot delta is the pips spot delta *premium-adjusted* by percentage foreign option value. This can be explained by assuming FOR is the numeraire. If one wants to hedge a short call of N notional in FOR with a premium of $NV_{fd}/\mathcal{X}$ in FOR, the delta sensitivity with respect to the spot inverse $\hat{\mathcal{X}}\equiv 1/\mathcal{X}$ must be

$$\frac{\partial \frac{V_{fd}}{\mathcal{X}}}{\partial \frac{1}{\mathcal{X}}} = \frac{\frac{1}{\mathcal{X}} \partial V_{fd} - \frac{V_{fd}}{\mathcal{X}^2} \partial \mathcal{X}}{-\frac{1}{\mathcal{X}^2} \partial \mathcal{X}} = V_{fd} - \mathcal{X}\Delta \tag{68}$$

To hedge the delta risk, one must be long $N\left(V_{fd} - \mathcal{X}\Delta\right)$ amount of the spot inverse $1/\mathcal{X}$. This can be achieved by entering a long position in $N\left(V_{fd} - \mathcal{X}\Delta\right)$ units of DOM with a cost of $N\left(V_{fd}/\mathcal{X} - \Delta\right)$ units of FOR. Or equivalently, one enters a long position in $N\left(\Delta - V_{fd}/\mathcal{X}\right)$ units of FOR with a cost of $N\left(\mathcal{X}\Delta - V_{fd}\right)$ units of DOM, which translates exactly into the percentage spot delta $\Delta_{\%} = \Delta - V_{\%f}$.

Whether pips or percentage delta is quoted in markets depends on which currency in the currency pair FORDOM is the premium currency, and the definition of premium currency itself is a market convention. If the premium currency is DOM, then no premium adjustment is applied and the pips delta is used, whereas if the premium currency is FOR then the percentage delta is used. Despite the fact that market convention involves different delta quotation styles, they are mutually equivalent to one another (referring to [13] for more details). The difference between pips delta and percentage delta comes naturally from the change of measure between domestic and foreign risk-neutral measures. Consider the case of a call option on FORDOM, or to be more thorough, a FOR call/DOM put. If the two counterparties to such a trade are FOR based and DOM based respectively, then they will agree on the price. However, the price will be expressed and actually exchanged in one of two currencies: FOR or DOM. From a domestic investor's point of view, if the premium currency is DOM, the premium itself is riskless and the hedging of the option can be done by simply taking Δ amount of FORDOM spot. If however the premium currency is FOR, there will be two sources of currency risk: 1) the change in intrinsic option value due to the move in underlying spot. 2) the change in premium value converted from FOR to DOM due to the move in FX rate. Apparently to hedge the two risks, one must take Δ and $-V_{fd}/\mathcal{X}_t$ amount of spot position respectively. Alternatively speaking, the *premium adjustment* comes from the fact that a premium in FOR would have already hedged part of the option's delta risk [14], which must be accounted in calculating the delta.

2.5.1.3. *Pips Forward Delta*

The pips forward delta is the ratio of the change in forward value (in contrast to present value!) of the option to the change in the relevant FX forward, both in domestic per foreign terms

$$\Delta_F = \frac{\partial V_{F;fd}}{\partial F} = \omega\Phi(\omega d_+) = \frac{\Delta}{\hat{P}} \tag{69}$$

by the following facts

$$V_{F;fd} = \frac{V_{fd}}{P} = \omega F\Phi(\omega d_+) - \omega K\Phi(\omega d_-), \qquad \frac{\partial d_+}{\partial F} = \frac{\partial d_-}{\partial F} = \frac{1}{F\sigma\sqrt{\tau}} \tag{70}$$

2.5.1.4. *Percentage Forward Delta*

The percentage forward delta is defined as the ratio of the change in forward value of the option to the change in the FX forward, both in percentage foreign terms

$$\Delta_{\%F} = \frac{\partial V_{F;\%f}}{\frac{\partial F}{F}} = F\frac{\partial}{\partial F}\left(\frac{V_{F;fd}}{F}\right) = \frac{\partial V_{F;fd}}{\partial F} - \frac{V_{F;fd}}{F} = \Delta_F - V_{F;\%f} = \omega\frac{K}{F}\Phi(\omega d_-) \tag{71}$$

Again, the percentage forward delta is the pips forward delta premium-adjusted by forward percentage foreign option value.

The choice between spot delta and forward delta depends on the currency pair as well as the option maturity. Spot delta is mainly used for tenors less than or equal to 1Y and for the currency pair with both currencies from the more developed economies. Otherwise, the use of forward delta dominates. It is obvious that the spot delta and forward delta differ only by a foreign discount factor $\hat{P}_{t,T}$. Since the credit crunch of 2008 and the associated low levels of liquidity in short-term interest rate products, it became unfeasible for banks to agree on spot deltas (which include discount factors) and, as a result, market practice has largely shifted to using forward deltas exclusively in the construction of the FX smile, which do not include any discounting [15].

2.5.1.5. *Strike from Delta Conversion*

It is straightforward to compute strikes from pips deltas. However, since explicit strike expressions in percentage deltas are not available, we must solve for the strikes numerically. It can be seen that the percentage deltas are monotonic in strike on put side, but this is not the case on call side. Using percentage forward delta as an example, the expression of a call delta is

$$\begin{aligned}\Delta_{\%F} &= \frac{K}{F}\Phi\left(\frac{1}{\sigma\sqrt{\tau}}\log\frac{F}{K}-\frac{\sigma\sqrt{\tau}}{2}\right)\\ \Delta_{\%F} &= -\frac{K}{F}\Phi\left(-\frac{1}{\sigma\sqrt{\tau}}\log\frac{F}{K}+\frac{\sigma\sqrt{\tau}}{2}\right)\end{aligned} \tag{72}$$

Obviously, the delta has two sources of dependence on strike and the function is not always monotonic. This may result in two different solutions of strike. To avoid the undesired solution, the numerical search can be performed within a range (K_{min}, K_{max}) that encloses the proper strike solution. We can choose the strike by pips delta as the upper bound K_{max} (because a pips delta maps to a strike that is always larger than that of a percentage delta) and the lower bound K_{min} can be found numerically as a solution to the equation below (where K_{min} maximizes the $\Delta_{\%F}$) [16]

$$\frac{\partial\Delta_{\%F}}{\partial K}=\frac{\Phi(d_-)}{F}-\frac{1}{F\sigma\sqrt{\tau}}\phi(d_-)=0 \Longrightarrow \Phi(d_-)\sigma\sqrt{\tau}=\phi(d_-) \tag{73}$$

However, the function below

$$f(K)=\Phi(d_-)\sigma\sqrt{\tau}-\phi(d_-) \tag{74}$$

is also not monotonic. It has a maximum $\sigma\sqrt{\tau}$ when $K\to 0$ and a minimum when $K=F\exp\left(\frac{1}{2}\sigma^2\tau\right)$, which can be used to find the K_{min}. The table below summarizes the delta and strike conversion of the 4 delta conventions.

Table 1. Deltas and delta neutral straddle strikes

Delta Convention	Delta from Strike	Strike from Delta
pips spot	$\Delta(K)=\omega\hat{P}\Phi(\omega d_+)$	$K(\delta\|\Delta)=F\exp\left(\frac{\sigma^2\tau}{2}-\omega\sigma\sqrt{\tau}\Phi^{-1}\left(\frac{\omega\delta}{\hat{P}}\right)\right)$
pips forward	$\Delta_F(K)=\omega\Phi(\omega d_+)$	$K(\delta\|\Delta_F)=F\exp\left(\frac{\sigma^2\tau}{2}-\omega\sigma\sqrt{\tau}\Phi^{-1}(\omega\delta)\right)$
percentage spot	$\Delta_{\%}(K)=\omega\hat{P}\frac{K}{F}\Phi(\omega d_-)$	$K(\delta\|\Delta_{\%})\in\left(K_{min},K(\delta\|\Delta)\right)$ for $\omega=1$

percentage forward	$\Delta_{\%F}(K) = \omega \dfrac{K}{F}\Phi(\omega d_-)$	$K(\delta\vert\Delta_{\%F}) \in \left(K_{min}, K(\delta\vert\Delta_F)\right)$ for $\omega = 1$

2.5.2. Other Risk Sensitivities

In the following context, we will only express the risk sensitivities in domestic per foreign terms for simplicity. Assuming the present value or the undiscounted of an option is given in the Black-Scholes model, e.g., $V_{fd} = P\mathfrak{B}(\omega, K, \sigma, \tau)$ or $V_{F;fd} = \mathfrak{B}(\omega, K, \sigma, \tau)$, in domestic currency, the risk sensitivities can be derived as follows.

2.5.2.1. *Theta*

Theta θ is the first derivative of the option price with respect to the initial time t. Converting from t to τ, we have $\theta = \partial V/\partial t = -\partial V/\partial \tau$. Let's first derive the partial derivatives

$$\begin{aligned}
\frac{\partial d_+}{\partial \tau} &= \frac{\partial\left(\left(\frac{\mu}{\sigma}+\frac{\sigma}{2}\right)\sqrt{\tau} + \frac{1}{\sigma\sqrt{\tau}}\log\frac{\mathcal{X}_t}{K}\right)}{\partial \tau} = \frac{\mu}{2\sigma\sqrt{\tau}} + \frac{\sigma}{4\sqrt{\tau}} - \frac{1}{2\sigma\sqrt{\tau^3}}\log\frac{\mathcal{X}}{K} \\
\frac{\partial d_-}{\partial \tau} &= \frac{\partial\left(d_+ - \sigma\sqrt{\tau}\right)}{\partial \tau} = \frac{\mu}{2\sigma\sqrt{\tau}} - \frac{\sigma}{4\sqrt{\tau}} - \frac{1}{2\sigma\sqrt{\tau^3}}\log\frac{\mathcal{X}}{K}
\end{aligned} \tag{75}$$

The theta can then be derived as

$$\begin{aligned}
\theta = \frac{\partial V}{\partial t} &= \omega\hat{r}\hat{P}\mathcal{X}\Phi(\omega d_+) - \hat{P}\mathcal{X}\phi(d_+)\frac{\partial d_+}{\partial \tau} - \omega r P K\Phi(\omega d_-) + PK\phi(d_-)\frac{\partial d_-}{\partial \tau} \\
&= \omega\hat{r}\hat{P}\mathcal{X}\Phi(\omega d_+) - \omega r P K\Phi(\omega d_-) - \hat{P}\mathcal{X}\phi(d_+)\frac{\sigma}{2\sqrt{\tau}}
\end{aligned} \tag{76}$$

where we have used the identity $\hat{P}\mathcal{X}\phi(d_+) = PK\phi(d_-) \implies F\phi(d_+) = K\phi(d_-)$.

2.5.2.2. *Gamma*

Spot (forward) Gamma Γ is the first derivative of the spot (forward) delta Δ with respect to the underlying spot $\mathcal{X}_t$ (forward $F_{t,T}$), or equivalently the second derivative of the present (undiscounted) value of the option with respect to the spot (forward)

$$\Gamma = \frac{\partial^2 V}{\partial \mathcal{X}^2} = \frac{\partial \Delta}{\partial \mathcal{X}} = \frac{\hat{P}\phi(d_+)}{\mathcal{X}\sigma\sqrt{\tau}}, \qquad \Gamma_F = \frac{\partial^2 V_F}{\partial F^2} = \frac{\partial \Delta_F}{\partial F} = \frac{\phi(d_+)}{F\sigma\sqrt{\tau}} \tag{77}$$

The call and the put option with an equal strike have the same gamma sensitivity.

2.5.2.3. *Vega*

Vega $\mathcal{V}$ is the first derivative of the option price with respect to the volatility σ, that is

$$\begin{aligned}
\mathcal{V} &= \frac{\partial V}{\partial \sigma} = \hat{P}\mathcal{X}\phi(d_+)\frac{\partial d_+}{\partial \sigma} - PK\phi(d_-)\frac{\partial d_-}{\partial \sigma} = \hat{P}\mathcal{X}\phi(d_+)\frac{d_+ - d_-}{\sigma} = \hat{P}\mathcal{X}\phi(d_+)\sqrt{\tau} \\
&= PK\phi(d_-)\sqrt{\tau} \\
\mathcal{V}_F &= \frac{\partial V_F}{\partial \sigma} = \frac{\partial}{\partial \sigma}\big(F\Phi(d_+) - K\Phi(d_-)\big) = F\phi(d_+)\frac{\partial d_+}{\partial \sigma} - K\phi(d_-)\frac{\partial d_-}{\partial \sigma} = K\phi(d_-)\frac{d_+ - d_-}{\sigma} \\
&= F\phi(d_+)\sqrt{\tau} = K\phi(d_-)\sqrt{\tau}`
\end{aligned} \tag{78}$$

where we have used the following equations

$$\begin{aligned}
\frac{\partial d_+}{\partial \sigma} &= \frac{\partial\left(\frac{1}{\sigma\sqrt{\tau}}\log\frac{F}{K} + \frac{\sigma\sqrt{\tau}}{2}\right)}{\partial \sigma} = -\frac{1}{\sigma^2\sqrt{\tau}}\log\frac{F}{K} + \frac{\sqrt{\tau}}{2} = -\frac{d_+}{\sigma} + \sqrt{\tau} = -\frac{d_-}{\sigma} \\
\frac{\partial d_-}{\partial \sigma} &= \frac{\partial\big(d_+ - \sigma\sqrt{\tau}\big)}{\partial \sigma} = \frac{\partial d_+}{\partial \sigma} - \sqrt{\tau} = -\frac{d_+}{\sigma}
\end{aligned} \tag{79}$$

The call and the put option with an equal strike have the same vega sensitivity.

2.5.2.4. *Vanna*

Vanna is the cross derivative of the present (undiscounted) option value with respect to the spot (forward) and the volatility σ. The Vanna can be derived as

$$\begin{aligned}
\frac{\partial^2 V}{\partial \mathcal{X}\partial \sigma} &= \frac{\partial \Delta}{\partial \sigma} = \hat{P}\phi(d_+)\frac{\partial d_+}{\partial \sigma} = -\frac{\hat{P}\phi(d_+)d_-}{\sigma} = -\frac{\mathcal{V}d_-}{\mathcal{X}\sigma\sqrt{\tau}} \\
\frac{\partial^2 V}{\partial \mathcal{X}\partial \sigma} &= \frac{\partial \mathcal{V}}{\partial \mathcal{X}} = -PK\phi(d_-)\sqrt{\tau}d_-\frac{\partial d_-}{\partial \mathcal{X}} = -\frac{\mathcal{V}d_-}{\mathcal{X}\sigma\sqrt{\tau}} \\
\frac{\partial^2 V_F}{\partial F\partial \sigma} &= \frac{\partial \Delta_F}{\partial F} = \frac{\partial \Phi(d_+)}{\partial \sigma} = -\frac{\phi(d_+)d_-}{\sigma}
\end{aligned} \tag{80}$$

$$\frac{\partial^2 V_F}{\partial F \partial \sigma} = \frac{\partial \mathcal{V}_F}{\partial F} = -K\phi(d_-)\sqrt{\tau}d_- \frac{\partial d_+}{\partial F} = -\mathcal{V}_F d_- \frac{\partial d_+}{\partial F} = -\frac{\mathcal{V}_F d_-}{F\sigma\sqrt{\tau}} = -\frac{\phi(d_+)d_-}{\sigma}$$

using the facts

$$\frac{\partial d_+}{\partial \mathcal{X}} = \frac{\partial d_-}{\partial \mathcal{X}} = \frac{1}{\mathcal{X}\sigma\sqrt{\tau}}, \qquad \frac{\partial d_+}{\partial F} = \frac{\partial d_-}{\partial F} = \frac{1}{F\sigma\sqrt{\tau}} \tag{81}$$

The call and the put option with the same strike have the same vanna sensitivity.

2.5.2.5. *Volga*

Volga is the second derivative of the option price with respect to the volatility σ

$$\begin{aligned} \frac{\partial^2 V}{\partial \sigma^2} &= \frac{\partial \mathcal{V}}{\partial \sigma} = \hat{P}\mathcal{X}\sqrt{\tau}\frac{\partial \phi(d_+)}{\partial d_+}\frac{\partial d_+}{\partial \sigma} = \hat{P}\mathcal{X}\sqrt{\tau}\phi(d_+)\frac{d_+ d_-}{\sigma} = \frac{\mathcal{V} d_+ d_-}{\sigma} \\ \frac{\partial^2 V_F}{\partial \sigma^2} &= \frac{\partial \mathcal{V}_F}{\partial \sigma} = -F\phi(d_+)d_+\sqrt{\tau}\frac{\partial d_+}{\partial \sigma} = F\phi(d_+)d_+\sqrt{\tau}\frac{d_-}{\sigma} = \frac{\mathcal{V}_F\, d_+ d_-}{\sigma} \end{aligned} \tag{82}$$

using the fact that

$$\frac{\partial \phi(d_+)}{\partial d_+} = \frac{\partial\left(\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{d_+^2}{2}\right)\right)}{\partial d_+} = -\phi(d_+)d_+ \tag{83}$$

The call and the put option with an equal strike have the same volga sensitivity.

2.6. FX Volatility Smile Convention

In liquid FX markets, Straddle, Risk Reversal and Butterfly are some of the most traded option strategies. It is convention that the markets usually quote volatilities instead of the direct prices of these instruments, and typically express these volatilities as functions of delta, e.g., $\delta = 0.25$ or 0.1, which are commonly referred to as the 25-Delta or the 10-Delta. Let's define a general form of delta function $\Delta(\omega, K, \sigma)$, whick can be any of the pips spot Δ, pips forward Δ_F, percentage spot $\Delta_\%$ or percentage forward $\Delta_{\%F}$. The δ in Black-Scholes model can be computed by the delta function $\Delta(\omega, K, \sigma)$ from a strike K and a volatility σ. Providing a market consistent volatility smile $\sigma(K)$ at a maturity, there is a 1-to-1 mapping from δ to K such that $\delta = \Delta\big(\omega, K, \sigma(K)\big)$.

2.6.1. At-The-Money Volatility

FX markets quote the at-the-money volatility σ_{atm} against a conventionally defined at-the-money strike K_{atm}. There are mainly two types of at-the-money definitions: ATM forward (ATMFWD) and ATM delta-neutral straddle (ATMDNS). A market consistent volatility smile $\sigma(K)$ must admit the fact that $\sigma(K_{atm}) = \sigma_{atm}$.

2.6.1.1. *ATM Forward*

In this definition, the at-the-money strike is set to the FX forward $F_{t,T}$

$$K_{atm} = F_{t,T} \tag{84}$$

This convention is used for currency pairs including a Latin American emerging market currency, e.g., MXN, BRL, etc. It may also apply to options with maturities longer than 10Y.

2.6.1.2. *Delta Neutral Straddle*

A delta-neutral straddle (DNS) is a straddle with zero combined call and put delta, such as

$$\Delta(1, K_{atm}, \sigma_{atm}) + \Delta(-1, K_{atm}, \sigma_{atm}) = 0 \tag{85}$$

If the $\Delta(\omega, K, \sigma)$ is in the form of pips spot delta (65) or pips forward delta (69), the ATM strike K_{atm} corresponding to the ATM volatility σ_{atm} can be derived as

$$\Phi(d_+) - \Phi(-d_+) = 0 \Longrightarrow \Phi(d_+) = 0.5 \Longrightarrow K_{atm} = F \exp\left(\frac{1}{2}\sigma_{atm}^2 \tau\right) \tag{86}$$

Alternatively, if the $\Delta(K, \sigma, \omega)$ takes the form of percentage spot delta (67) or percentage forward delta (71), the ATM strike K_{atm} can be derived as

$$\Phi(d_-) - \Phi(-d_-) = 0 \Longrightarrow \Phi(d_-) = 0.5 \Longrightarrow K_{atm} = F \exp\left(-\frac{1}{2}\sigma_{atm}^2 \tau\right) \tag{87}$$

The table below summarizes the ATMFWD and ATMDNS strikes with associated delta definitions

Table 2. Deltas and delta neutral straddle strikes

Delta Convention	Delta Formula	ATMFWD Delta	ATMDNS Strike	ATMDNS Delta
pips spot	$\omega\hat{P}\Phi(\omega d_+)$	$\omega\hat{P}\Phi\left(\omega\frac{\sigma_{atm}\sqrt{\tau}}{2}\right)$	$F\exp\left(\frac{\sigma_{atm}^2\tau}{2}\right)$	$\frac{1}{2}\omega\hat{P}$
pips forward	$\omega\Phi(\omega d_+)$	$\omega\Phi\left(\omega\frac{\sigma_{atm}\sqrt{\tau}}{2}\right)$	$F\exp\left(\frac{\sigma_{atm}^2\tau}{2}\right)$	$\frac{1}{2}\omega$

percentage spot	$\omega \hat{P} \dfrac{K}{F} \Phi(\omega d_-)$	$\omega \hat{P} \Phi\left(-\omega \dfrac{\sigma_{atm}\sqrt{\tau}}{2}\right)$	$F \exp\left(-\dfrac{\sigma_{atm}^2 \tau}{2}\right)$	$\dfrac{1}{2} \omega \hat{P} \exp\left(-\dfrac{\sigma_{atm}^2 \tau}{2}\right)$
percentage forward	$\omega \dfrac{K}{F} \Phi(\omega d_-)$	$\omega \Phi\left(-\omega \dfrac{\sigma_{atm}\sqrt{\tau}}{2}\right)$	$F \exp\left(-\dfrac{\sigma_{atm}^2 \tau}{2}\right)$	$\dfrac{1}{2} \omega \exp\left(-\dfrac{\sigma_{atm}^2 \tau}{2}\right)$

It is evident that if the ATM strike is greater (smaller) than the forward, the market convention must be that deltas for that currency pair are quoted as pips (percentage) deltas [17].

2.6.2. Risk Reversal Volatility

FX markets quote the risk reversal volatility $\sigma_{\delta RR}$ as a difference between the δ-delta call and put volatilities. Providing a market consistent volatility smile $\sigma(K)$, it is given by

$$\sigma_{\delta RR} = \sigma(K_{\delta C}) - \sigma(K_{\delta P}) \tag{88}$$

where δ-delta smile strikes $K_{\delta C}$ and $K_{\delta P}$ can be inverted from the delta function such that

$$\Delta\big(1, K_{\delta C}, \sigma(K_{\delta C})\big) = \delta, \qquad \Delta\big(-1, K_{\delta P}, \sigma(K_{\delta P})\big) = -\delta \tag{89}$$

2.6.3. Strangle Volatility

There are two types of strangle volatilities.

2.6.3.1. *Market Strangle*

Market strangle (MS, also known as brokers fly) is quoted as a single volatility $\sigma_{\delta MS}$ for a delta δ. The δ-delta market strangle strikes $K_{MS,\delta C}$ and $K_{MS,\delta P}$ for the call and put are both calculated in Black-Scholes model with a single constant volatility of $\sigma_{atm} + \sigma_{\delta MS}$, such that at these strikes the call and put have deltas of $\pm\delta$ respectively

$$\Delta\big(1, K_{MS,\delta C}, \sigma_{atm} + \sigma_{\delta MS}\big) = \delta, \qquad \Delta\big(-1, K_{MS,\delta P}, \sigma_{atm} + \sigma_{\delta MS}\big) = -\delta \tag{90}$$

This gives the value of the market strangle in Black-Sholes model as

$$V_{\delta MS} = \mathfrak{B}\big(1, K_{MS,\delta C}, \sigma_{atm} + \sigma_{\delta MS}, \tau\big) + \mathfrak{B}\big(-1, K_{MS,\delta P}, \sigma_{atm} + \sigma_{\delta MS}, \tau\big) \tag{91}$$

This value must be satisfied by a market consistent volatility smile $\sigma(K)$, such that the $V'_{\delta MS}$ defined below must be equal to the $V_{\delta MS}$

$$V'_{\delta MS} = \mathfrak{B}\big(1, K_{MS,\delta C}, \sigma(K_{MS,\delta C}), \tau\big) + \mathfrak{B}\big(-1, K_{MS,\delta P}, \sigma(K_{MS,\delta P}), \tau\big) \tag{92}$$

Note that, at these strikes we generally have

$$\Delta\left(1, K_{MS,\delta C}, \sigma\big(K_{MS,\delta C}\big)\right) \neq \delta, \qquad \Delta\left(-1, K_{MS,\delta P}, \sigma\big(K_{MS,\delta P}\big)\right) \neq -\delta \tag{93}$$

Providing a calibrated volatility smile $\sigma(K)$ consistent with the market, it is easy to derive the market strangle volatility from the smile. The procedure takes the following steps

1. Choose an initial guess for $\sigma_{\delta MS}$ (e.g., let $\sigma_{\delta MS} = \sigma_{\delta SS}$)
2. Compute the market strangle strikes $K_{MS,\delta C}$ and $K_{MS,\delta P}$ by inverting (90) given the δ
3. Compute the strangle value $V_{\delta MS}$ in (91) and the $V'_{\delta MS}$ in (92)
4. If $V'_{\delta MS}$ is close to $V_{\delta MS}$ then the $V_{\delta MS}$ is found, otherwise go to step 1 to repeat the iteration

2.6.3.2. *Smile Strangle*

Providing a market consistent volatility smile $\sigma(K)$ is available, it is more intuitive to express the strangle volatility $\sigma_{\delta SS}$ as

$$\sigma_{\delta SS} = \frac{\sigma(K_{\delta C}) + \sigma(K_{\delta P})}{2} - \sigma(K_{atm}) \tag{94}$$

This is called smile strangle volatility, where the smile strikes $K_{\delta C}$ and $K_{\delta P}$ are given by (89).

Given the market quoted σ_{atm}, $\sigma_{\delta RR}$ and $\sigma_{\delta MS}$, one can build a volatility smile $\sigma(K)$ that is consistent with the market. The procedure takes the following steps

1) Preparation:
   - Determine the delta convention (e.g., pips or percentage, spot or forward)
   - Determine the at-the-money convention (e.g., ATMFWD or ATMDNS) and its associated ATM strike K_{atm}
   - Choose a parametric form for the volatility smile $\sigma(K)$ (e.g., Polynomial-in-Delta interpolation)
   - Determine the market strangle strikes $K_{MS,\delta C}$ and $K_{MS,\delta P}$ by (90) using $\sigma_{atm} + \sigma_{\delta MS}$
   - Compute the value of market strangle $V_{\delta MS}$ in (91)

2) Choose an initial guess for $\sigma_{\delta SS}$ (e.g., $\sigma_{\delta SS} = \sigma_{\delta MS}$)

3) Use σ_{atm}, $\sigma_{\delta RR}$ and $\sigma_{\delta SS}$ to find the best fit of $\sigma(K)$ such that with the smile strikes $K_{\delta C}$ and $K_{\delta P}$ given by (89), we have

$$\begin{aligned} &\sigma(K_{atm}) = \sigma_{atm} \\ &\sigma(K_{\delta C}) - \sigma(K_{\delta P}) = \sigma_{\delta RR} \\ &\frac{\sigma(K_{\delta C}) + \sigma(K_{\delta P})}{2} - \sigma(K_{atm}) = \sigma_{\delta SS} \end{aligned} \tag{95}$$

4) Compute the value of the market strangle $V'_{\delta MS}$ in (92) with the market strangle strikes $K_{MS,\delta C}$ and $K_{MS,\delta P}$ using the $\sigma(K)$ fitted in step 3).

5) If $V'_{\delta MS}$ is close to the true market strangle $V_{\delta MS}$ then the $\sigma(K)$ is found, otherwise go to step 2) to repeat the iteration.

2.6.4. Smile Volatility

From the relationship in (95), we can easily find the implied volatilities corresponding to δ-delta smile strikes $K_{\delta C}$ and $K_{\delta P}$

$$\sigma(K_{\delta P}) = \sigma_{atm} + \sigma_{\delta SS} - \frac{\sigma_{\delta RR}}{2}, \qquad \sigma(K_{\delta C}) = \sigma_{atm} + \sigma_{\delta SS} + \frac{\sigma_{\delta RR}}{2} \tag{96}$$

where the δ-delta smile strikes $K_{\delta C}$ and $K_{\delta P}$ can be solved from (89).

3. VOLATILITY SURFACE CONSTRUCTION

Table 3 presents an example of ATM, risk reversal and smile strangle volatilities at a series of maturities. Each maturity may associate with different ATM and delta conventions. In previous section, we have shown how to extract the five volatilities, at $\pm 10D$ $\pm 25D$ and ATM respectively, from market quotes for each maturity subject to its associated market convention. It is often desired to have a volatility surface, so that an implied volatility at arbitrary delta/strike and maturity can be interpolated from the surface.

Table 3. ATM, risk reversal and smile strangle volatilities with associated conventions

Maturity	ATM Convention	Delta Convention	ST10D	ST25D	ATM	RR25D	RR10D
1M	ATMDNS	Spot Percentage	0.73%	0.28%	9.13%	-1.13%	-2.09%
3M	ATMDNS	Spot Percentage	1.01%	0.36%	9.59%	-1.43%	-2.72%
6M	ATMDNS	Spot Percentage	1.33%	0.44%	10.00%	-1.66%	-3.15%
1Y	ATMDNS	Spot Percentage	1.67%	0.51%	10.39%	-1.88%	-3.66%
3Y	ATMDNS	Forward Percentage	2.34%	0.68%	10.58%	-1.90%	-3.59%
5Y	ATMDNS	Forward Percentage	2.65%	0.74%	10.86%	-2.00%	-3.64%
7Y	ATMDNS	Forward Percentage	2.80%	0.73%	11.36%	-2.20%	-3.85%
10Y	ATMDNS	Forward Percentage	2.75%	0.57%	12.43%	-2.63%	-4.60%
12Y	ATMFWD	Forward Percentage	2.23%	0.64%	12.73%	-2.78%	-4.44%
15Y	ATMFWD	Forward Percentage	2.16%	0.62%	13.03%	-3.13%	-5.07%
20Y	ATMFWD	Forward Percentage	2.13%	0.63%	13.03%	-3.18%	-5.08%

3.1. Smile Interpolation

There are many ways to perform a smile interpolation, specifically to interpolate a 5-point volatility smile. We are going to introduce a few practical interpolation methods as follows.

3.1.1. Polynomial-in-Delta

Polynomial-in-Delta is one of the simple and widely used methods. It employs a 4th order polynomial which allows a perfect fit to five volatilities of a smile (or a 2nd order polynomial if just fitting to three volatilities. such 3-point fit has been introduced in [18]). The parameterization is as follows

$$\log \sigma(K) = \sum_{j=0}^{4} a_j x(K)^j , \qquad x(K) = M(K) - M(Z) \tag{97}$$

where a_j's are the coefficients to be calibrated (exactly) to the market volatilities. The function $M(\cdot)$ provides a measure of moneyness that often takes the form

$$M(K) = \Phi\left(\frac{1}{v\sqrt{\tau}}\log\frac{K}{\Lambda}\right) \tag{98}$$

where Λ can be the forward F or the at-the-money strike K_{atm}. Polynomial-in-Delta interpolation is named after the fact that the measure of moneyness (98) is similar to the definition of forward delta (69). The parameter Z in (97) can be chosen to be the F or the K_{atm} such that the $x(K)$ provides a measure of distance in moneyness from the Z. The parameter v in (98) is a normalizing volatility to be determined later.

Calibration of the coefficients a_j's is straightforward. From previous discussion, we are able to retrieve 5 volatility-strike pairs (σ_i, K_i) for $i = 1, \cdots, 5$ at a given maturity from market quotes, i.e., volatilities at strikes corresponding to $\pm 10D$, $\pm 25D$ and ATM subject to prevailing delta and ATM conventions. Based on the 5 volatilities, we are able to form a full rank linear system from (97), which can then be solved for the coefficients a_j's.

The parameter v in (98) need be determined. For simplicity, we may take $v = \sigma_{atm}$. To be more adaptive, one may choose $v = \sigma(K)$. This has no impact to the calibration. But we must then solve (97) iteratively to interpolate the volatility $\sigma(K)$. Using adaptive $\sigma(K)$ for v is usually desired for improving wing behavior of the smile, though computationally inefficient. To mitigate this issue, we may proceed with a prediction-correction scheme. In this scheme, we calibrate the model twice, one with $v = \sigma_{atm}$ which gives calibrated coefficients $\hat{a}_j$'s, and the other one with $v = \sigma(K)$ which gives calibrated coefficients a_j's. To find the interpolated volatility, we first use $\hat{a}_j$'s along with $v = \sigma_{atm}$ to obtain a prediction $\hat{\sigma}(K)$. This quantity should be a good approximation of the true value of $\sigma(K)$. We then use coefficients a_j's along with $v = \hat{\sigma}(K)$ to derive a further improved approximation of $\sigma(K)$. Numerical experiments confirm that the values derived from the prediction-correction scheme are in an excellent agreement with those computed from iterative solver.

3.1.2. Stochastic Volatility Inspired (SVI)

Stochastic Volatility Inspired (SVI) parameterization was introduced by Gatheral in 2004 [19] [20]. The idea is to build a smooth parameterization of the smile which guarantees a wing behavior, such that, the implied variance w is always linear in k, e.g., $w(k) \propto k$ for $k \to \pm\infty$, where w is the square of implied volatility and $k = \log K$ is the log-strike. Such wing behavior, which is consistent with stochastic volatility assumption, is backed by theoretical arguments that can be found in Lee [21]. The raw SVI parameterization reads

$$v_{BS}^2(k) = w(k; a, b, \rho, m, \sigma) = a + b\left(\rho(k-m) + \sqrt{(k-m)^2 + \sigma^2}\right) \tag{99}$$

When $\sigma = 0$, we can obtain the left and right asymptotes

$$\begin{aligned} w_L(k; a, b, \rho, m, \sigma) &= a - b(1-\rho)(k-m) \\ w_R(k; a, b, \rho, m, \sigma) &= a + b(1+\rho)(k-m) \end{aligned} \tag{100}$$

It follows immediately that changes in the parameters have the following effects

- a gives the overall level of variance. Increasing a increases the overall level of variance, a vertical translation of the smile. The overall level must be somewhat bounded by the largest observed total variance, hence we have $a \le \max\{w_i\}$
- b gives the angle between the left and right asymptotes. Increasing b increases the slopes of both the put and call wings, tightening the smile. Since volatility smiles usually have positive ATM curvature, it indicates $b \ge 0$. A necessary condition for the absence of dynamic arbitrage, gives an upper bound, $b \le 4/(1+|\rho|)$
- ρ determines the orientation of the graph. Increasing ρ decreases (increases) the slope of the left (right) wing, a counter-clockwise rotation of the smile. since ρ is basically a factor that explains the correlation between the spot and the volatility process, we need to have $-1 \le \rho \le 1$
- m translates the graph. Increasing m translates the smile to the right. Since m is a quantity associated with k, we may define the range $\min\{k_i\} \le m \le \max\{k_i\}$

- σ determines how smooth the vertex is. Increasing σ reduces the at-the-money (ATM) curvature of the smile. In general, $\sigma > 0$ as it is an empirical fact that volatility smiles have a positive at-the-money curvature. No upper bound can be derived for σ, but a small integer always does the job, for example $\sigma < 10$

We want to calibrate the model (99) to market quoted 5 variance/log-strike pairs (w_i, k_i) for $i = 1, \cdots, 5$ at a given maturity, i.e., implied variance at log strikes corresponding to $\pm 10D$, $\pm 25D$ and ATM subject to proper delta and ATM conventions. This is a 5-dimensional nonlinear root finding problem that finds parameters $\{a, b, \rho, m, \sigma\}$ such that

$$v_{BS}^2(k_i) = w(k_i; a, b, \rho, m, \sigma) \quad \text{for} \quad i = 1, \cdots, 5 \tag{101}$$

Or equivalently, we may find an optimal set of parameters $\{a, b, \rho, m, \sigma\}$ such that the objective function

$$f(a, b, \rho, m, \sigma) = \sum_{i=1}^{5} \left(v_{BS}^2(k_i) - w(k_i; a, b, \rho, m, \sigma) \right)^2 \tag{102}$$

is minimized, which is basically a nonlinear least square problem. However, calibration of this model can be difficult numerically [22] [23] [24] because of the high dimensionality of the problem (5 parameters) and also because the parameters are not completely "orthogonal" (e.g., varying m and ρ both change the skewness; varying b and σ both change the convexity). As such, the objective function (102) usually has multiple local minima which renders gradient methods unreliable.

3.1.2.1. *Uni-SVI*

In order to calibrate the model efficiently, we transform the raw SVI model (99) equivalently to

$$w(k; \alpha, \beta, m, \gamma, \sigma) = \alpha + \beta(k - m) + \gamma \left(\sqrt{(k - m)^2 + \sigma^2} - \sigma \right) \tag{103}$$

by employing the change of variables $\rho = \beta / \gamma$, $b = \gamma$ and $a = \alpha - \gamma\sigma$. We call this model *Uni-SVI*. In this form of SVI model

- α gives the overall level of volatility smile
- β determines the skewness of volatility smile

- m translates the volatility smile
- γ controls the convexity of volatility smile
- σ determines smoothness of the vertex around $k = m$

We want to calibrate the model to the 5 variance/log-strike pairs $\{(w_i, k_i)|\ i = 10p,\ 25p,\ atm,\ 25c,\ 10c\}$ corresponding to the 5 different delta values. For a fast and reliable calibration, a good initial guess must be provided. In the model, we first suggest an intuitive initial guess for parameter m with $\widehat{m} = k_{atm}$ and for σ with $\hat{\sigma} = \left(k_{25c} - k_{25p}\right)/2$. This gives the equation

$$w(k) = \alpha + \beta(k - \widehat{m}) + \gamma\left(\sqrt{(k - \widehat{m})^2 + \hat{\sigma}^2} - \hat{\sigma}\right) \tag{104}$$

When $k = k_{atm}$, it is easy to see that $\hat{\alpha} = w_{atm}$. Since $\left|k_{10p} - k_{atm}\right|$ and $|k_{10c} - k_{atm}|$ are usually more than twice as large as $\hat{\sigma}$, when $k = k_{10p}$ we may take an approximation and write

$$\begin{aligned} k = k_{10p} \implies w_{10p} &\approx \hat{\alpha} + \hat{\beta}\left(k_{10p} - \widehat{m}\right) - \hat{\gamma}\left(k_{10p} - \widehat{m}\right) - \hat{\gamma}\hat{\sigma} \\ \implies \hat{\beta} - \hat{\gamma}\phi_p &= \frac{w_{atm} - w_{10p}}{k_{atm} - k_{10p}} = \delta_p, \qquad \phi_p = 1 - \frac{\hat{\sigma}}{k_{atm} - k_{10p}} \end{aligned} \tag{105}$$

Similarly, when $k = k_{10c}$ we write

$$\begin{aligned} k = k_{10c} \implies w_{10c} &= \hat{\alpha} + \hat{\beta}(k_{10c} - \widehat{m}) + \hat{\gamma}(k_{10c} - \widehat{m}) - \hat{\gamma}\hat{\sigma} \\ \implies \hat{\beta} + \hat{\gamma}\phi_c &= \frac{w_{10c} - w_{atm}}{k_{10c} - k_{atm}} = \delta_c, \qquad \phi_c = 1 - \frac{\hat{\sigma}}{k_{10c} - k_{atm}} \end{aligned} \tag{106}$$

Combining (105) and (106), we end up with the following initial guess for the parameters

$$\hat{\alpha} = w_{atm}, \qquad \hat{\beta} = \delta_p + \hat{\gamma}\phi_p, \qquad \hat{\gamma} = \frac{\delta_c - \delta_p}{\phi_c + \phi_p}, \qquad \widehat{m} = k_{atm}, \qquad \hat{\sigma} = \frac{k_{25c} - k_{25p}}{2} \tag{107}$$

Numerical experiments show that this initial guess generally leads to a fast convergence of a nonlinear solver even without imposing parameter bounds.

3.1.2.2. *Bi-SVI*

Available upon request …

3.1.2.3. *Tri-SVI*

Available upon request …

3.2. Temporal Interpolation

The most commonly used temporal interpolation assumes a flat forward volatility in time. This is equivalent to a linear interpolation in total variance. For example, if we have $\sigma_{atm}(p)$ and $\sigma_{atm}(q)$ at maturities p and q respectively, subject to the same ATM and delta convention, we may interpolate an ATM volatility at a time t for $p < t < q$ by the formula

$$\sigma_{atm}^2(t)t = \frac{q-t}{q-p}\sigma_{atm}^2(p)p + \frac{t-p}{q-p}\sigma_{atm}^2(q)q \tag{108}$$

The temporal interpolation in $\pm 10D$ and $\pm 25D$ volatilities are in the same manner.

3.3. Volatility Surface by Standard Conventions

Table 2 shows that the market convention on ATM and delta style may vary from one maturity to another. Such jumps in conventions introduce inconsistency in definition of the ATM strikes and δ-deltas across maturities. It will be much convenient to assume a unified standard convention for marking ATM strike and δ-deltas strikes at all maturities [25]. A pragmatic choice is to use delta-neutral-straddle ATM and forward pips delta as the *standard* convention. For each maturity t, we convert the 5-point volatility-strike pairs (σ_i, K_i) associated with a specific market convention to $\left(\tilde{\sigma}_i, \widetilde{K}_i\right)$ such that the new 5-point $\left(\tilde{\sigma}_i, \widetilde{K}_i\right)$ pairs conform to the standard ATM and delta convention. This conversion involves first building a smile using the 5-point (σ_i, K_i) pairs with the market convention, and then finding from the smile the 5-point $\left(\tilde{\sigma}_i, \widetilde{K}_i\right)$ pairs with the standard convention, for $i = 10p, 25p, atm, 25c, 10c$.

To find a smile at an interim time u for $p < u < q$ between two adjacent maturities p and q, the temporal interpolation is performed on the volatilities with standard convention. For example, we get a $25c$ volatility $\tilde{\sigma}_{25c}(u)$ at the interim time u by interpolating from $\tilde{\sigma}_{25c}(p)$ and $\tilde{\sigma}_{25c}(q)$. We obtain all the 5-point volatilities $\tilde{\sigma}_i(u)$, along with their associated strikes $\widetilde{K}_i(u)$ (which are inverted from the δ-delta values given the standard ATM and delta convention). The last step is then to use the 5-point volatility-strike pairs $\left(\tilde{\sigma}_i(u), \widetilde{K}_i(u)\right)$ to build a volatility smile, again with standard ATM and delta convention, for strike interpolation at time u.

4. THE VANNA-VOLGA METHOD

The vanna-volga method is a technique for pricing first-generation FX exotic products (e.g., barriers, digitals and touches, etc.). The main idea of vanna-volga method is to adjust the Black-Scholes theoretical value (TV) of an option by adding the smile cost of a portfolio that hedges three main risks associated to the volatility of the option: the vega, vanna and volga.

4.1. Vanna-Volga Pricing

Suppose there exists a portfolio H with a long position in an exotic trade Y, a short position in Δ amount of the underlying spot $\mathcal{X}$, and short positions in ω_1 amount of instrument A_1, ω_2 amount of instrument A_2 and ω_3 amount of instrument A_3. The hedging instruments A_i's can be the straddle, risk reversal and butterfly, as they are liquidly traded in FX markets and they carry *mainly* vega, vanna and volga risks respectively that can be used to hedge the volatility risks of the trade Y. By construction, the price of the portfolio and its dynamics must follow

$$H = Y - \Delta\mathcal{X} - \sum_{i=1}^{3}\omega_i A_i\,, \qquad dH = dY - \Delta d\mathcal{X} - \sum_{i=1}^{3}\omega_i dA_i \tag{109}$$

We may estimate the Greeks in Black-Scholes model and further express the price dynamics in terms of the stochastic spot $\mathcal{X}$ and flat volatility σ. By Ito's lemma, we have

$$\begin{aligned} dH = &\underbrace{\left(\frac{\partial Y}{\partial t} - \sum_{i=1}^{3}\omega_i\frac{\partial A_i}{\partial t}\right)}_{\text{Theta}} dt + \underbrace{\left(\frac{\partial Y}{\partial \mathcal{X}} - \Delta - \sum_{i=1}^{3}\omega_i\frac{\partial A_i}{\partial \mathcal{X}}\right)}_{\text{Delta}} d\mathcal{X} \\ &+ \frac{1}{2}\underbrace{\left(\frac{\partial^2 Y}{\partial \mathcal{X}^2} - \sum_{i=1}^{3}\omega_i\frac{\partial^2 A_i}{\partial \mathcal{X}^2}\right)}_{\text{Gamma}} d\mathcal{X}d\mathcal{X} \\ &+ \underbrace{\left(\frac{\partial Y}{\partial \sigma} - \sum_{i=1}^{3}\omega_i\frac{\partial A_i}{\partial \sigma}\right)}_{\text{Vega}} d\sigma + \frac{1}{2}\underbrace{\left(\frac{\partial^2 Y}{\partial \sigma^2} - \sum_{i=1}^{3}\omega_i\frac{\partial^2 A_i}{\partial \sigma^2}\right)}_{\text{Volga}} d\sigma d\sigma + \underbrace{\left(\frac{\partial^2 Y}{\partial \mathcal{X}\partial\sigma} - \sum_{i=1}^{3}\omega_i\frac{\partial^2 A_i}{\partial \mathcal{X}\partial\sigma}\right)}_{\text{Vanna}} d\mathcal{X}d\sigma \end{aligned} \tag{110}$$

Choosing the Δ and the weights ω_i so as to zero out the coefficients of $d\mathcal{X}$, $d\sigma$, $d\sigma d\sigma$ and $d\mathcal{X}d\sigma$, the portfolio is then *locally* risk free at time t (given that the gamma and other higher order risks can be ignored) and must have a return at risk free rate. Therefore, when the flat volatility is stochastic and the options are valued in Black-Scholes model, we can still have a *locally* perfect hedge. The perfect hedge in the three volatility risks implies that the following linear system must be satisfied

$$\begin{pmatrix} \text{vega}(Y) \\ \text{vanna}(Y) \\ \text{volga}(Y) \end{pmatrix} = \begin{pmatrix} \text{vega}(A_1) & \text{vega}(A_2) & \text{vega}(A_3) \\ \text{vanna}(A_1) & \text{vanna}(A_2) & \text{vanna}(A_3) \\ \text{volga}(A_1) & \text{volga}(A_2) & \text{volga}(A_3) \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \tag{111}$$

This perfect hedging is under an assumption of flat volatility. Due to non-flat nature of the volatility surface, additional cost between $A_i(\sigma_{\text{smile}})$ and $A_i(\sigma_{\text{flat}})$ must be accounted into the price of the trade Y to fulfil the hedging. As a result, the vanna-volga price Y_{VV} of the trade Y is computed as follows

$$Y_{VV}(\sigma_{\text{smile}}) = Y_{TV}(\sigma_{\text{flat}}) + \sum_{i=1}^{3} \omega_i \big(A_i(\sigma_{\text{smile}}) - A_i(\sigma_{\text{flat}})\big) \tag{112}$$

where $Y_{TV}(\sigma_{\text{flat}})$ is the theoretical Black-Scholes value using a flat volatility (e.g., $\sigma_{\text{flat}} = \sigma_{atm}$), $A_i(\sigma_{\text{smile}})$ and $A_i(\sigma_{\text{flat}})$ are the prices of the hedging instrument valued with a volatility smile and a flat volatility respectively.

4.2. Smile Interpolation

The vanna-volga method may also serve a purpose of interpolating a volatility smile based on the market quoted at-the-money volatility σ_{atm}, the δ-delta risk reversal volatility $\sigma_{\delta RR}$, and lastly the δ-delta smile strangle volatility $\sigma_{\delta SS}$ (converted from market strangle volatility $\sigma_{\delta MS}$ by the method in section 2.6.3.2). From the relationship in (95), we can derive the following quantities

Strikes	Implied Volatilities
$K_1 = K_{\delta P}$	$\sigma_1 = \sigma(K_{\delta P}) = \sigma_{atm} + \sigma_{\delta SS} - \dfrac{\sigma_{\delta RR}}{2}$
$K_2 = K_{atm}$	$\sigma_2 = \sigma(K_{atm}) = \sigma_{atm}$
$K_3 = K_{\delta C}$	$\sigma_3 = \sigma(K_{\delta C}) = \sigma_{atm} + \sigma_{\delta SS} + \dfrac{\sigma_{\delta RR}}{2}$

where the ATM strike K_{atm} is given by the at-the-money convention, and the δ-delta smile strikes $K_{\delta C}$ and $K_{\delta P}$ are solved from (89).

We will follow a similar analysis in section 4.1. Suppose we have a perfect hedged portfolio P that consists of a long position in a call option Y with an arbitrary strike K, a short position in Δ amount of spot $\mathcal{X}$, and three short positions in ω_i amount of call options A_i with strikes $K_1 = K_{\delta P}$, $K_2 = K_{atm}$ and $K_3 = K_{\delta C}$. The perfect hedge in the three volatility risks admits that the following linear system must be satisfied

$$\begin{pmatrix} \mathrm{vega}(Y) \\ \mathrm{vanna}(Y) \\ \mathrm{volga}(Y) \end{pmatrix} = \begin{pmatrix} \mathrm{vega}(A_1) & \mathrm{vega}(A_2) & \mathrm{vega}(A_3) \\ \mathrm{vanna}(A_1) & \mathrm{vanna}(A_2) & \mathrm{vanna}(A_3) \\ \mathrm{volga}(A_1) & \mathrm{volga}(A_2) & \mathrm{volga}(A_3) \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \tag{113}$$

where these volatility sensitivities can be estimated in Black-Scholes model assuming a flat volatility flat volatility σ (usually we choose $\sigma = \sigma_{atm}$). Plugging the closed form Black-Scholes vega, vanna and volga in (78) (80) and (82) respectively, the (113) becomes

$$\mathcal{V}(K) \begin{pmatrix} 1 \\ d_+ d_-(K) \\ d_-(K) \end{pmatrix} = \begin{pmatrix} \mathcal{V}(K_1) & \mathcal{V}(K_2) & \mathcal{V}(K_3) \\ \mathcal{V} d_+ d_-(K_1) & \mathcal{V} d_+ d_-(K_2) & \mathcal{V} d_+ d_-(K_3) \\ \mathcal{V} d_-(K_1) & \mathcal{V} d_-(K_2) & \mathcal{V} d_-(K_3) \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \tag{114}$$

where $d_+ d_-(K)$ is short for $d_+(K) d_-(K)$ and $\mathcal{V} d_+ d_-(K)$ for $\mathcal{V}(K) d_+(K) d_-(K)$. By inverting the linear system, there is a unique solution of ω for the strike K, such that

$$\omega_1 = \frac{\mathcal{V}(K)}{\mathcal{V}(K_1)} \frac{\log\frac{K_2}{K}\log\frac{K_3}{K}}{\log\frac{K_2}{K_1}\log\frac{K_3}{K_1}}, \qquad \omega_2 = \frac{\mathcal{V}(K)}{\mathcal{V}(K_2)} \frac{\log\frac{K}{K_1}\log\frac{K_3}{K}}{\log\frac{K_2}{K_1}\log\frac{K_3}{K_2}}, \qquad \omega_3 = \frac{\mathcal{V}(K)}{\mathcal{V}(K_3)} \frac{\log\frac{K}{K_1}\log\frac{K}{K_2}}{\log\frac{K_3}{K_1}\log\frac{K_3}{K_2}} \tag{115}$$

A “smile-consistent” volatility v (i.e., a Black Scholes volatility implied from the price by the vanna-volga method) for the call with the strike K is then obtained by adding to the Black-Scholes price the cost of implementing the above hedging at prevailing market prices, that is

$$C(K, v) = C(K, \sigma) + \sum_{i=1}^{3} \omega_i \big(C(K_i, \sigma_i) - C(K_i, \sigma) \big) \tag{116}$$

where the function $C(K,\sigma)$ stands for the Black-Scholes call option price with strike K and flat volatility σ.

A market implied volatility curve can then be constructed by inverting (116), for each considered K. Here we introduce an approximation approach. By taking the first order expansion of (116) in σ, that is we approximate $C(K_i,\sigma_i) - C(K_i,\sigma)$ by $(\sigma_i - \sigma)\mathcal{V}(K_i)$, we have

$$C(K,v) \approx C(K,\sigma) + \sum_{i=1}^{3} \omega_i (\sigma_i - \sigma)\mathcal{V}(K_i) \tag{117}$$

Substituting ω_i with the results in (115) and using the fact that $\mathcal{V}(K) = \sum_{i=1}^{3} \omega_i \mathcal{V}(K_i)$, we have

$$C(K,v) \approx C(K,\sigma) + \mathcal{V}(K)\left(\sum_{\mathrm{i}=1}^{3} y_i\sigma_i - \sigma\right) \approx C(K,\sigma) + \mathcal{V}(K)(\bar{v} - \sigma) \Longrightarrow \bar{v} \approx \sum_{\mathrm{i}=1}^{3} y_i\sigma_i \tag{118}$$

where $\bar{v}$ is the first order approximation of the implied volatility v for strike K, and the coefficients y_i are given by

$$y_1 = \frac{\log\frac{K_2}{K}\log\frac{K_3}{K}}{\log\frac{K_2}{K_1}\log\frac{K_3}{K_1}}, \qquad y_2 = \frac{\log\frac{K}{K_1}\log\frac{K_3}{K}}{\log\frac{K_2}{K_1}\log\frac{K_3}{K_2}}, \qquad y_3 = \frac{\log\frac{K}{K_1}\log\frac{K}{K_2}}{\log\frac{K_3}{K_1}\log\frac{K_3}{K_2}} \tag{119}$$

This shows that the implied volatility v can be approximated by a linear combination of the three smile volatilities σ_i.

A more accurate second order approximation, which is asymptotically constant at extreme strikes, can be obtained by expanding the (116) at second order in σ

$$\begin{aligned} C(K,v) &\approx C(K,\sigma) + \mathcal{V}(K)(\bar{\bar{v}} - \sigma) + \frac{1}{2}\mathcal{V}_\sigma(K)(\bar{\bar{v}} - \sigma)^2 \\ &\approx C(K,\sigma) + \sum_{\mathrm{i}=1}^{3} \omega_i \left(\mathcal{V}(K_i)(\sigma_i - \sigma) + \frac{1}{2}\mathcal{V}_\sigma(K_i)(\sigma_i - \sigma)^2\right) \end{aligned} \tag{120}$$

$$\Longrightarrow \mathcal{V}(K)(\bar{\bar{v}} - \sigma) + \frac{\mathcal{V}d_+d_-(K)}{2\sigma}(\bar{\bar{v}} - \sigma)^2$$

$$\approx \mathcal{V}(K)\sum_{\mathrm{i=1}}^{3} y_i\sigma_i - \mathcal{V}(K)\sigma + \frac{\mathcal{V}(K)}{2\sigma}\sum_{\mathrm{i=1}}^{3} y_i d_+d_-(K_i)(\sigma_i - \sigma)^2$$

$$\Longrightarrow \frac{d_+d_-(K)}{2\sigma}(\bar{\bar{v}} - \sigma)^2 + (\bar{\bar{v}} - \sigma) - \left(\bar{v} - \sigma + \frac{\sum_{i=1}^{3} y_i d_+d_-(K_i)(\sigma_i - \sigma)^2}{2\sigma}\right) \approx 0$$

Solving the quadratic equation in (120) gives the second order approximation

$$\bar{\bar{v}} \approx \sigma + \frac{-\sigma + \sqrt{\sigma^2 + \left(2\sigma(\bar{v} - \sigma) + \sum_{i=1}^{3} y_i d_+d_-(K_i)(\sigma_i - \sigma)^2\right)d_+d_-(K)}}{d_+d_-(K)} \tag{121}$$

where $d_+d_-(K)$ is evaluated with a flat volatility σ.

5. CLASSIC LOCAL VOLATILITY: DUPIRE

In local volatility models, the volatility process is assumed to be a function of time and FX (or equity) spot level. It is one step generalization of the well-known Black-Scholes model. In the following, we are going to introduce Dupire local volatility, which as mentioned in Section 1.3 can be regarded as the conditional risk-neutral expectation of the instantaneous future variance (i.e., conditional mean of the stochastic volatility). To show this, we may assume under risk neutral measure the FX spot process $\mathcal{X}_t$ follows a general SDE below

$$\frac{d\mathcal{X}_t}{\mathcal{X}_t} = \mu_t dt + \sigma_t dW_t, \qquad \mu_t = r_t - \hat{r}_t \tag{122}$$

with domestic short rate r_t and foreign short rate $\hat{r}_t$ (or dividend rate for equity). Usually, we use a "hat" accent to denote quantities in foreign economy. The volatility process σ_t is a general function of time. It can be stochastic and may also be dependent on spot level $\mathcal{X}_t$. In the context of the Dupire local volatility model, the volatility process is simplified to be a deterministic function of the spot $\mathcal{X}_t$, such that $\sigma_t = \ell(t, \mathcal{X}_t)$.

5.1. Local Volatility by Vanilla Call

The price of European (vanilla) call option at initial time s can be expressed as a function of running maturity t and strike K

$$\mathcal{C}_{t,K|s,\alpha} = \mathbb{E}_s\left[D_{s,t}(\mathcal{X}_t - K)^+\right] = P_{s,t}\mathbb{E}_s^t[(\mathcal{X}_t - K)^+] = P_{s,t}\int_K^\infty (x - K)p_{t,x|s,\alpha}dx \tag{123}$$

where $\mathbb{E}_s^t[\cdot]$ denotes an expectation under t-forward measure. The value of domestic zero coupon bond is given by

$$P_{s,t} = \mathbb{E}_s\left[\exp\left(-\int_s^t r_u du\right)\right] = \exp\left(-\int_s^t f_{s,u} du\right) \tag{124}$$

where $f_{s,t}$ is the forward rate of r_t (i.e., the deterministic interest rate). The $p_{t,x|s,\alpha}$, which is under t-forward measure, is the transition probability density having spot $\mathcal{X}_t = x$ at t given initial condition $\mathcal{X}_s = \alpha$ at s. For brevity, we will use $\mathcal{C}_{t,K}$ for call price, short for $\mathcal{C}_{t,K|s,\alpha}$, and $p_{t,x}$ short for $p_{t,x|s,\alpha}$.

Under assumption of deterministic r_t and $\hat{r}_t$, we may write $r_t = f_{s,t}$ and $\hat{r}_t = \hat{f}_{s,t}$. Differentiating (123) with respect to K, we have the first order and second order partial derivatives

$$\frac{\partial \mathcal{C}_{t,K}}{\partial K} = -P_{s,t}\int_K^\infty p_{t,x}dx\,, \qquad \frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2} = P_{s,t}p_{t,K} \tag{125}$$

The cumulative density function and the probability density function of the transition probability can then be expressed as

$$\int_{-\infty}^K p_{t,x}dx = 1 - \frac{1}{P_{s,t}}\frac{\partial \mathcal{C}_{t,K}}{\partial K}, \qquad p_{t,K} = \frac{1}{P_{s,t}}\frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2} \tag{126}$$

The (126) is also known as *Breeden-Litzenberger* formula.

Taking partial derivative of $\mathcal{C}_{t,K}$ with respect to t, we find that

$$\begin{aligned}\frac{\partial \mathcal{C}_{t,K}}{\partial t} &= -r_t\mathcal{C}_{t,K} + P_{s,t}\int_K^\infty (x-K)\frac{\partial p_{t,x}}{\partial t}dx \\ &= -r_t\mathcal{C}_{t,K} + P_{s,t}\int_K^\infty (x-K)\left(\frac{1}{2}\frac{\partial^2\left(\ell_{t,x}^2 x^2 p_{t,x}\right)}{\partial x^2} - \frac{\partial\left(\mu_t x p_{t,x}\right)}{\partial x}\right)dx\end{aligned} \tag{127}$$

where we have used the *Kolmogorov Forward* Equation (11)

$$\frac{\partial p_{t,x}}{\partial t} = \frac{1}{2}\frac{\partial^2\left(\ell_{t,x}^2 x^2 p_{t,x}\right)}{\partial x^2} - \frac{\partial\left(\mu_t x p_{t,x}\right)}{\partial x} \tag{128}$$

Applying integration by parts to the integrals on the right hand side of (127) yields

$$\begin{aligned}\int_K^\infty (x-K)\frac{\partial\left(\mu_t x p_{t,x}\right)}{\partial x}dx &= \underbrace{(x-K)\mu_t x p_{t,x}\Big|_{x=K}^{\infty}}_{=0} - \mu_t\int_K^\infty x p_{t,x}dx \\ &= -\mu_t K\int_K^\infty p_{t,x}dx - \mu_t\int_K^\infty (x-K)p_{t,x}dx = \frac{\mu_t K}{P_{s,t}}\frac{\partial \mathcal{C}_{t,K}}{\partial K} - \frac{\mu_t \mathcal{C}_{t,K}}{P_{s,t}}\end{aligned} \tag{129}$$

and

$$\int_K^\infty (x-K)\frac{\partial^2\left(\ell_{t,x}^2 x^2 p_{t,x}\right)}{\partial x^2}dx = \underbrace{(x-K)\frac{\partial\left(\ell_{t,x}^2 x^2 p_{t,x}\right)}{\partial x}\bigg|_{x=K}^{\infty}}_{=0} - \int_K^\infty \frac{\partial\left(\ell_{t,x}^2 x^2 p_{t,x}\right)}{\partial x}dx$$
$$= -\ell_{t,x}^2 x^2 p_{t,x}\Big|_{x=K}^{\infty} = \ell_{t,K}^2 K^2 p_{t,K} = \frac{\ell_{t,K}^2 K^2}{P_{s,t}}\frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2} \tag{130}$$

where we have $\lim_{x\to\infty} p_{t,x} = 0$ and $\lim_{x\to\infty} \partial p_{t,x}/\partial x = 0$ assuming that the density function $p_{t,x}$ and its first derivative vanish at a higher order of rate as $x \to \infty$. Plugging (129) and (130) into (127), we find that

$$\frac{\partial \mathcal{C}_{t,K}}{\partial t} = \frac{\ell_{t,K}^2 K^2}{2}\frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2} - \mu_t K\frac{\partial \mathcal{C}_{t,K}}{\partial K} - \hat{r}_t \mathcal{C}_{t,K} \tag{131}$$

and eventually we reach the classic Dupire formula for the local volatility $\ell_{t,K}$

$$\ell_{t,K}^2 = \frac{\dfrac{\partial \mathcal{C}_{t,K}}{\partial t} + \mu_t K\dfrac{\partial \mathcal{C}_{t,K}}{\partial K} + \hat{r}_t \mathcal{C}_{t,K}}{\dfrac{1}{2}K^2\dfrac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2}} \tag{132}$$

The classic Dupire formula expressed in put options can be derived in the same manner. Alternatively, one can quickly obtain its expression from (132) by put-call parity relation $\mathcal{C}_{t,K} = \mathcal{P}_{t,K} + P_{s,t}\left(F_{s,t} - K\right)$

$$\ell_{t,K}^2 = \frac{\dfrac{\partial \mathcal{P}}{\partial t} - r_t P_{s,t}\left(F_{s,t} - K\right) + \mu_t P_{s,t} F_{s,t} + \mu_t K\left(\dfrac{\partial \mathcal{P}}{\partial K} - P_{s,t}\right) + \hat{r}_t\left(\mathcal{P} + P_{s,t}\left(F_{s,t} - K\right)\right)}{\dfrac{1}{2}K^2\dfrac{\partial^2 \mathcal{P}}{\partial K^2}}$$
$$= \frac{\dfrac{\partial \mathcal{P}_{t,K}}{\partial t} + \mu_t K\dfrac{\partial \mathcal{P}_{t,K}}{\partial K} + \hat{r}_t \mathcal{P}_{t,K}}{\dfrac{1}{2}K^2\dfrac{\partial^2 \mathcal{P}_{t,K}}{\partial K^2}} \tag{133}$$

where the FX forward is given as

$$F_{s,t} = \mathcal{X}_s \frac{\hat{P}_{s,t}}{P_{s,t}} \tag{134}$$

Numerical methods often demand a local volatility function constructed on a 2D grid, one dimension for time and the other for spot (or strike). It is often numerically more stable to work with spatial dimension in terms of log-strike or log-moneyness. For example, if we express the strike in log-moneyness k, the change of variable will be

$$k = \log \frac{K}{F_{s,t}} \quad \text{and} \quad \frac{\partial k}{\partial K} = \frac{1}{K}, \qquad \frac{\partial k}{\partial t} = -\mu_t \tag{135}$$

where the forward is given by

$$F_{s,t} = \mathcal{X}_s \exp\left(\int_s^t \mu_u du\right), \qquad \frac{\partial F_{s,t}}{\partial t} = \mu_t F_{s,t} \tag{136}$$

We want to express the classic Dupire formula in the (t, k)-plane, using call $\mathcal{C}_{t,k} = \mathcal{C}_{t,k|s,\alpha}$, which is equivalent to the call $\mathcal{C}_{t,K} = \mathcal{C}_{t,K|s,\alpha}$. The transformation from (t, K) to (t, k) is achieved by using the following partial derivatives derived by chain rule

$$\begin{aligned}
\frac{\partial \mathcal{C}_{t,K}}{\partial t} &= \frac{\partial \mathcal{C}_{t,k}}{\partial t} + \frac{\partial \mathcal{C}_{t,k}}{\partial k}\frac{\partial k}{\partial t} = \frac{\partial \mathcal{C}_{t,k}}{\partial t} - \mu_t \frac{\partial \mathcal{C}_{t,k}}{\partial k} \\
\frac{\partial \mathcal{C}_{t,K}}{\partial K} &= \frac{\partial \mathcal{C}_{t,k}}{\partial t}\frac{\partial t}{\partial K} + \frac{\partial \mathcal{C}_{t,k}}{\partial k}\frac{\partial k}{\partial K} = \frac{1}{K}\frac{\partial \mathcal{C}_{t,k}}{\partial k} \\
\frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2} &= \frac{\partial}{\partial t}\left(\frac{\partial \mathcal{C}_{t,k}}{\partial k}\right)\frac{\partial t}{\partial K} + \frac{\partial}{\partial k}\left(\frac{\partial \mathcal{C}_{t,k}}{\partial k}\right)\frac{\partial k}{\partial K} = \frac{1}{K}\frac{\partial}{\partial k}\left(\frac{1}{K}\frac{\partial \mathcal{C}_{t,k}}{\partial k}\right) = \frac{1}{K^2}\left(\frac{\partial^2 \mathcal{C}_{t,k}}{\partial k^2} - \frac{\partial \mathcal{C}_{t,k}}{\partial k}\right)
\end{aligned} \tag{137}$$

Plugging these partial derivatives into (132), we have the classic Dupire formula expressed in k as

$$\ell_{t,k}^2 = \frac{\dfrac{\partial \mathcal{C}_{t,k}}{\partial t} + \hat{r}_t \mathcal{C}_{t,k}}{\dfrac{1}{2}\left(\dfrac{\partial^2 \mathcal{C}_{t,k}}{\partial k^2} - \dfrac{\partial \mathcal{C}_{t,k}}{\partial k}\right)} \tag{138}$$

5.2. Local Volatility by Undiscounted Call

Given deterministic rates, it sometimes is more convenient to express the classic Dupire formula in terms of an undiscounted call value $C_{t,K} = C_{t,K|s,\alpha}$, that is

$$C_{t,K|s,\alpha} = \mathbb{E}_s[(\mathcal{X}_t - K)^+] = \int_K^\infty (x - K) p_{t,x|s,\alpha} dx = \frac{\mathcal{C}_{t,K|s,\alpha}}{P_{s,t}} \tag{139}$$

with the discounted call $\mathcal{C}_{t,K|s,\alpha}$ defined in (123). We can derive the partial derivatives of the undiscounted call, similar to those in (125) and (131)

$$\frac{\partial C_{t,K}}{\partial K} = -\int_K^\infty p_{t,x|s,\alpha} dx\,, \qquad \frac{\partial^2 C_{t,K}}{\partial K^2} = p_{t,K|s,\alpha} \tag{140}$$

$$\frac{\partial C_{t,K}}{\partial t}=\int_K^\infty (x-K)\frac{\partial p_{t,x|s,\alpha}}{\partial t}dx=\frac{1}{2}\ell_{t,K}^2K^2\frac{\partial^2 C_{t,K}}{\partial K^2}+\mu_t\left(C_{t,K}-K\frac{\partial C_{t,K}}{\partial K}\right)$$

The transition probability cumulative density function has a simple expression as

$$\int_{-\infty}^K p_{t,x}dx=1+\frac{\partial C_{t,K}}{\partial K} \tag{141}$$

which allows us to estimate the cumulative density numerically using a call spread[1]. The Dupire formula for $\ell_{t,K}$ in the undiscounted call hence reads

$$\ell_{t,K}^2=\frac{\frac{\partial C_{t,K}}{\partial t}+\mu_t K\frac{\partial C_{t,K}}{\partial K}-\mu_t C_{t,K}}{\frac{1}{2}K^2\frac{\partial^2 C_{t,K}}{\partial K^2}} \tag{142}$$

Using log-moneyness k for the strike, we again want to derive the Dupire formula in (t,k) using undiscounted call $C_{t,k}$, which is equivalent to the undiscounted call $C_{t,K}$. The transformation from (t,K) to (t,k) is done through using the following partial derivatives given by chain rule

$$\begin{aligned}
\frac{\partial C_{t,K}}{\partial t}&=\frac{\partial C_{t,k}}{\partial t}+\frac{\partial C_{t,k}}{\partial k}\frac{\partial k}{\partial t}=\frac{\partial C_{t,k}}{\partial t}-\mu_t\frac{\partial C_{t,k}}{\partial k}\\
\frac{\partial C_{t,K}}{\partial K}&=\frac{\partial C_{t,k}}{\partial t}\frac{\partial t}{\partial K}+\frac{\partial C_{t,k}}{\partial k}\frac{\partial k}{\partial K}=\frac{1}{K}\frac{\partial C_{t,k}}{\partial k}\\
\frac{\partial^2 C_{t,K}}{\partial K^2}&=\frac{\partial}{\partial t}\left(\frac{\partial C_{t,k}}{\partial k}\right)\frac{\partial t}{\partial K}+\frac{\partial}{\partial k}\left(\frac{\partial C_{t,k}}{\partial k}\right)\frac{\partial k}{\partial K}=\frac{1}{K}\frac{\partial}{\partial k}\left(\frac{1}{K}\frac{\partial C_{t,k}}{\partial k}\right)=\frac{1}{K^2}\left(\frac{\partial^2 C_{t,k}}{\partial k^2}-\frac{\partial C_{t,k}}{\partial k}\right)
\end{aligned} \tag{143}$$

Plugging these partial derivatives into (142), we have the Dupire formula as follows

$$\frac{\ell_{t,k}^2}{2}=\frac{\frac{\partial C_{t,k}}{\partial t}-\mu_t C_{t,k}}{\frac{\partial^2 C_{t,k}}{\partial k^2}-\frac{\partial C_{t,k}}{\partial k}} \tag{144}$$

where $\ell_{t,k}$ is the local volatility in (t,k) equivalent to $\ell_{t,K}$.

5.3. Local Volatility by Implied Volatility

[1] When the $C_{t,K}$ is given as (undiscounted) Black-Scholes option value C_{BS}, we would have the cumulative density as $\int_{-\infty}^K p_{t,x}dx=1+\frac{\partial C_{BS}}{\partial K}+\frac{\partial C_{BS}}{\partial \sigma}\frac{\partial \sigma}{\partial K}=1-\Phi(d_-)+K\phi(d_-)\sqrt{\tau}\frac{\partial\sigma}{\partial K}=1-\Phi(d_-)+F\phi(d_+)\sqrt{\tau}\frac{\partial \sigma}{\partial K}$

Markets quote option prices as Black-Scholes implied volatilities. Hence, it is more straightforward to express the local volatility in terms of the implied volatilities rather than option prices. Taking s as the initial time, the undiscounted call price $B_{t,K,\xi}$ in Black-Scholes model is given by

$$B_{t,K,\xi} = P_{s,t}\left(F_{s,t}\Phi(d_+) - K\Phi(d_-)\right), \qquad d_\pm = \frac{1}{\xi_{t,K}\sqrt{\tau}}\log\frac{F_{s,t}}{K} \pm \frac{\xi_{t,K}\sqrt{\tau}}{2}, \qquad \tau = t - s \tag{145}$$

where $\xi = \xi_{t,K}$ is the Black-Scholes implied volatility from market quotes and Φ the standard normal cumulative density function (with density function ϕ). Its partial derivatives can be derived as

$$\begin{aligned}\frac{\partial B_{t,K,\xi}}{\partial t} &= -r_t B_{t,K,\xi} + \mu_t P_{s,t}F_{s,t}\Phi(d_+) + P_{s,t}F_{s,t}\phi(d_+)\frac{\partial d_+}{\partial t} - P_{s,t}K\phi(d_-)\frac{\partial d_-}{\partial t}\\ &= -r_t B_{t,K,\xi} + \mu_t P_{s,t}F_{s,t}\Phi(d_+) + P_{s,t}F_{s,t}\phi(d_+)\frac{\xi}{2\sqrt{\tau}}\end{aligned}$$

$$\begin{aligned}\frac{\partial B_{t,K,\xi}}{\partial K} &= P_{s,t}\left(F_{s,t}\phi(d_+)\frac{\partial d_+}{\partial K} - K\phi(d_-)\frac{\partial d_-}{\partial K} - \Phi(d_-)\right) = -P_{s,t}\Phi(d_-)\\ \frac{\partial B_{t,K,\xi}}{\partial \xi} &= P_{s,t}\left(F_{s,t}\phi(d_+)\frac{\partial d_+}{\partial \xi} - K\phi(d_-)\frac{\partial d_-}{\partial \xi}\right) = P_{s,t}F_{s,t}\phi(d_+)\sqrt{\tau} = P_{s,t}K\sqrt{\tau}\phi(d_-)\\ \frac{\partial^2 B_{t,K,\xi}}{\partial K^2} &= \frac{P_{s,t}\phi(d_-)}{K\xi\sqrt{\tau}}, \qquad \frac{\partial^2 B_{t,K,\xi}}{\partial \xi^2} = \frac{P_{s,t}d_+d_-K\phi(d_-)\sqrt{\tau}}{\xi}, \qquad \frac{\partial^2 B_{t,K,\xi}}{\partial \xi\partial K} = \frac{P_{s,t}\phi(d_-)d_+}{\xi}\end{aligned} \tag{146}$$

where we have used

$$\frac{\partial d_\pm}{\partial t} = \frac{\mu_t}{\xi\sqrt{\tau}} - \frac{d_\mp}{2\tau}, \qquad \frac{\partial d_\pm}{\partial K} = -\frac{1}{K\xi\sqrt{\tau}}, \qquad \frac{\partial d_\pm}{\partial \xi} = -\frac{d_\mp}{\xi}, \qquad \frac{\partial^2 d_\pm}{\partial K\partial \xi} = \frac{1}{K\xi^2\sqrt{\tau}} \tag{147}$$

and the identity

$$F_{s,t}\phi(d_+) = K\phi(d_-) \tag{148}$$

Further using the partial derivatives

$$\begin{aligned}&\frac{\partial C_{t,K}}{\partial t} = \frac{\partial B_{t,K,\xi}}{\partial t} + \frac{\partial B_{t,K,\xi}}{\partial \xi}\frac{\partial \xi}{\partial t}, \qquad \frac{\partial C_{t,K}}{\partial K} = \frac{\partial B_{t,K,\xi}}{\partial K} + \frac{\partial B_{t,K,\xi}}{\partial \xi}\frac{\partial \xi}{\partial K}\\ &\frac{\partial^2 C_{t,K}}{\partial K^2} = \frac{\partial^2 B_{t,K,\xi}}{\partial K^2} + 2\frac{\partial^2 B_{t,K,\xi}}{\partial K\partial \xi}\frac{\partial \xi}{\partial K} + \frac{\partial B_{t,K,\xi}}{\partial \xi}\frac{\partial^2 \xi}{\partial K^2} + \frac{\partial^2 B_{t,K,\xi}}{\partial \xi^2}\left(\frac{\partial \xi}{\partial K}\right)^2\end{aligned} \tag{149}$$

we can derive from (132) the local volatility expressed in implied volatility

$$\ell_{t,K}^2 = \frac{\frac{\partial B_{t,K,\xi}}{\partial t} + \frac{\partial B_{t,K,\xi}}{\partial \xi}\frac{\partial \xi}{\partial t} + \mu_t K \frac{\partial B_{t,K,\xi}}{\partial K} + \mu_t K \frac{\partial B_{t,K,\xi}}{\partial \xi}\frac{\partial \xi}{\partial K} + \hat{r}_t B_{t,K,\xi}}{\frac{1}{2}K^2\left(\frac{\partial^2 B_{t,K,\xi}}{\partial K^2} + 2\frac{\partial^2 B_{t,K,\xi}}{\partial K \partial \xi}\frac{\partial \xi}{\partial K} + \frac{\partial B_{t,K,\xi}}{\partial \xi}\frac{\partial^2 \xi}{\partial K^2} + \frac{\partial^2 B_{t,K,\xi}}{\partial \xi^2}\left(\frac{\partial \xi}{\partial K}\right)^2\right)}$$
$$= \frac{\xi^2 + 2\xi\tau\left(\frac{\partial \xi}{\partial t} + \mu_t K \frac{\partial \xi}{\partial K}\right)}{1 + 2\sqrt{\tau}K d_+ \frac{\partial \xi}{\partial K} + d_+ d_- \tau K^2 \left(\frac{\partial \xi}{\partial K}\right)^2 + \xi\tau K^2 \frac{\partial^2 \xi}{\partial K^2}} \tag{150}$$

5.3.1. Formula in Log Spot Moneyness

The local volatility formula (150) can also be expressed in log spot moneyness or log forward moneyness. Suppose we change the strike to the log spot moneyness $\mathscr{k} = \log(K/\mathcal{X}_s)$ for the implied volatility $\xi = \xi_{t,\mathscr{k}}$, the change of variable gives a local volatility expression as

$$\ell_{t,\mathscr{k}}^2 = \frac{\xi^2 + 2\xi\tau\left(\frac{\partial \xi}{\partial t} + \mu_t \frac{\partial \xi}{\partial \mathscr{k}}\right)}{1 + 2\sqrt{\tau} d_+ \frac{\partial \xi}{\partial \mathscr{k}} + d_+ d_- \tau \left(\frac{\partial \xi}{\partial \mathscr{k}}\right)^2 + \xi\tau\left(\frac{\partial^2 \xi}{\partial \mathscr{k}^2} - \frac{\partial \xi}{\partial \mathscr{k}}\right)}$$
$$= \frac{\xi^2 + 2\xi\tau\left(\frac{\partial \xi}{\partial t} + \mu_t \frac{\partial \xi}{\partial \mathscr{k}}\right)}{1 + \left(\xi\tau - 2\frac{k}{\xi}\right)\frac{\partial \xi}{\partial \mathscr{k}} + \left(\frac{k^2}{\xi^2} - \frac{\xi^2\tau^2}{4}\right)\left(\frac{\partial \xi}{\partial \mathscr{k}}\right)^2 + \xi\tau\left(\frac{\partial^2 \xi}{\partial \mathscr{k}^2} - \frac{\partial \xi}{\partial \mathscr{k}}\right)} \tag{151}$$
$$= \frac{\xi^2 + 2\xi\tau\left(\frac{\partial \xi}{\partial t} + \mu_t \frac{\partial \xi}{\partial \mathscr{k}}\right)}{\left(1 - \frac{k}{\xi}\frac{\partial \xi}{\partial \mathscr{k}}\right)^2 - \left(\frac{\xi\tau}{2}\frac{\partial \xi}{\partial \mathscr{k}}\right)^2 + \xi\tau\frac{\partial^2 \xi}{\partial \mathscr{k}^2}}$$

providing the following identities

$$\frac{\partial \xi}{\partial K} = \frac{\partial \xi}{\partial \mathscr{k}}\frac{\partial \mathscr{k}}{\partial K} = \frac{1}{K}\frac{\partial \xi}{\partial \mathscr{k}}, \qquad \frac{\partial^2 \xi}{\partial K^2} = \frac{\partial}{\partial \mathscr{k}}\left(\frac{1}{K}\frac{\partial \xi}{\partial \mathscr{k}}\right)\frac{\partial \mathscr{k}}{\partial K} = \frac{1}{K^2}\left(\frac{\partial^2 \xi}{\partial \mathscr{k}^2} - \frac{\partial \xi}{\partial \mathscr{k}}\right), \qquad \frac{\partial \mathscr{k}}{\partial K} = \frac{1}{K}$$
$$k = \log\frac{K}{F_{s,t}}, \qquad d_\pm = \frac{-k}{\xi\sqrt{\tau}} \pm \frac{\xi\sqrt{\tau}}{2} \tag{152}$$

5.3.2. Formula in Log Forward Moneyness

We may also want to have strike in log forward moneyness $k = \log\left(K/F_{s,t}\right)$, which is defined in (135), and express the local volatility in terms of the total implied variance $v = v_{t,k}$, which is equivalent to $\xi_{t,K}^2\tau$. We first derive the call value $B = B_{t,k,v}$ in Black-Scholes model, equivalent to (145)

$$B_{t,k,v} = P_{s,t}F_{s,t}\left(\Phi(d_+) - e^k\Phi(d_-)\right), \qquad d_\pm = \frac{-k}{\sqrt{v}} \pm \frac{\sqrt{v}}{2} \tag{153}$$

Its partial derivatives can be derived as

$$\begin{aligned}
\frac{\partial B_{t,k,v}}{\partial v} &= P_{s,t}F_{s,t}\left(\phi(d_+)\frac{\partial d_+}{\partial v} - e^k\phi(d_-)\frac{\partial d_-}{\partial v}\right) = \frac{P_{s,t}F_{s,t}\phi(d_+)}{2\sqrt{v}} = \frac{P_{s,t}K\phi(d_-)}{2\sqrt{v}} \\
\frac{\partial^2 B_{t,k,v}}{\partial v^2} &= \frac{\partial B_{t,k,v}}{\partial v}\left(-\frac{1}{2v} - d_+\frac{\partial d_+}{\partial v}\right) = \frac{\partial B_{t,k,v}}{\partial v}\left(\frac{k^2}{2v^2} - \frac{1}{2v} - \frac{1}{8}\right) \\
\frac{\partial B_{t,k,v}}{\partial k} &= P_{s,t}F_{s,t}\left(\phi(d_+)\frac{\partial d_+}{\partial k} - e^k\Phi(d_-) - e^k\phi(d_-)\frac{\partial d_-}{\partial k}\right) = -P_{s,t}F_{s,t}e^k\Phi(d_-) \\
\frac{\partial^2 B_{t,k,v}}{\partial k^2} &= -P_{s,t}F_{s,t}e^k\Phi(d_-) + P_{s,t}F_{s,t}e^k\phi(d_-)\frac{1}{\sqrt{v}} = \frac{\partial B_{t,k,v}}{\partial k} + 2\frac{\partial B_{t,k,v}}{\partial v} \\
\frac{\partial^2 B_{t,k,v}}{\partial k \partial v} &= \frac{\partial}{\partial k}\left(\frac{P_{s,t}F_{s,t}\phi(d_+)}{2\sqrt{v}}\right) = \frac{\partial B_{t,k,v}}{\partial v}(-d_+)\frac{\partial d_+}{\partial k} = \frac{\partial B_{t,k,v}}{\partial v}\left(\frac{1}{2} - \frac{k}{v}\right) \\
\frac{\partial B_{t,k,v}}{\partial t} &= -r_t B_{t,k,v} + P_{s,t}\left(\Phi(d_+) - e^k\Phi(d_-)\right)\frac{\partial F_{s,t}}{\partial t} = -\hat{r}_t B_{t,k,v}
\end{aligned} \tag{154}$$

providing the identities

$$\frac{\partial d_\pm}{\partial v} = \frac{k}{2\sqrt{v^3}} \pm \frac{1}{4\sqrt{v}}, \qquad \frac{\partial d_\pm}{\partial k} = -\frac{1}{\sqrt{v}} \tag{155}$$

Knowing (138), we can establish the connection of the local volatility $\ell_{t,k}$ to the total implied variance $v_{t,k}$ by deriving the following partial derivatives using chain rule

$$\begin{aligned}
&\frac{\partial C_{t,k}}{\partial t} = \frac{\partial B_{t,k,v}}{\partial t} + \frac{\partial B_{t,k,v}}{\partial v}\frac{\partial v}{\partial t}, \qquad \frac{\partial C_{t,k}}{\partial k} = \frac{\partial B_{t,k,v}}{\partial k} + \frac{\partial B_{t,k,v}}{\partial v}\frac{\partial v}{\partial k} \\
&\frac{\partial^2 C_{t,k}}{\partial k^2} = \frac{\partial}{\partial k}\frac{\partial C_{t,k}}{\partial k} + \frac{\partial}{\partial v}\frac{\partial C_{t,k}}{\partial k}\frac{\partial v}{\partial k} = \frac{\partial^2 B_{t,k,v}}{\partial k^2} + 2\frac{\partial^2 B_{t,k,v}}{\partial k \partial v}\frac{\partial v}{\partial k} + \frac{\partial B_{t,k,v}}{\partial v}\frac{\partial^2 v}{\partial k^2} + \frac{\partial^2 B_{t,k,v}}{\partial v^2}\left(\frac{\partial v}{\partial k}\right)^2
\end{aligned} \tag{156}$$

This gives the expression of local volatility as

$$\ell_{t,k}^2 = \frac{\frac{\partial \mathcal{C}_{t,k}}{\partial t} + \hat{r}_t \mathcal{C}_{t,k}}{\frac{1}{2}\left(\frac{\partial^2 \mathcal{C}_{t,k}}{\partial k^2} - \frac{\partial \mathcal{C}_{t,k}}{\partial k}\right)} = \frac{\frac{\partial B}{\partial t} + \frac{\partial B}{\partial v}\frac{\partial v}{\partial t} + \hat{r}_t B}{\frac{1}{2}\left(\frac{\partial^2 B}{\partial k^2} + 2\frac{\partial^2 B}{\partial k \partial v}\frac{\partial v}{\partial k} + \frac{\partial B}{\partial v}\frac{\partial^2 v}{\partial k^2} + \frac{\partial^2 B}{\partial v^2}\left(\frac{\partial v}{\partial k}\right)^2 - \frac{\partial B}{\partial k} - \frac{\partial B}{\partial v}\frac{\partial v}{\partial k}\right)}$$

$$= \frac{-\hat{r}_t B + \frac{\partial B}{\partial v}\frac{\partial v}{\partial t} + \hat{r}_t B}{\frac{1}{2}\left(\frac{\partial B}{\partial k} + 2\frac{\partial B}{\partial v} + 2\frac{\partial B}{\partial v}\left(\frac{1}{2} - \frac{k}{v}\right)\frac{\partial v}{\partial k} + \frac{\partial B}{\partial v}\frac{\partial^2 v}{\partial k^2} + \frac{\partial B}{\partial v}\left(\frac{k^2}{2v^2} - \frac{1}{2v} - \frac{1}{8}\right)\left(\frac{\partial v}{\partial k}\right)^2 - \frac{\partial B}{\partial k} - \frac{\partial B}{\partial v}\frac{\partial v}{\partial k}\right)} \quad (157)$$

$$= \frac{\frac{\partial v}{\partial t}}{1 - \frac{k}{v}\frac{\partial v}{\partial k} + \frac{1}{4}\left(\frac{k^2}{v^2} - \frac{1}{v} - \frac{1}{4}\right)\left(\frac{\partial v}{\partial k}\right)^2 + \frac{1}{2}\frac{\partial^2 v}{\partial k^2}} = \frac{\frac{\partial v}{\partial t}}{\left(\frac{k}{2v}\frac{\partial v}{\partial k} - 1\right)^2 - \left(\frac{1}{v} + \frac{1}{4}\right)\left(\frac{1}{2}\frac{\partial v}{\partial k}\right)^2 + \frac{1}{2}\frac{\partial^2 v}{\partial k^2}}$$

with the partial derivatives in (154).

5.3.3. Conversion between Log Forward Moneyness and Log Spot Moneyness

The expression for $\ell_{\mathcal{X}}(t, \Bbbk)$ in (151) and for $\ell_F(t, k)$ in (157) are often used to estimate the local volatilities on a temporal-spatial 2D grid. The local volatility $\ell_{\mathcal{X}}(t, \Bbbk)$ in log spot moneyness $\Bbbk = \log(K/\mathcal{X}_s)$ can be interpolated from a local volatility surface $\ell_F(t, k)$ in log forward moneyness $k = \log\left(K/F_{s,t}\right)$ through the following conversion, or vice-versa

$$\ell_{\mathcal{X}}(t, \Bbbk) = \ell_F(t, k) = \ell_F\left(t, \log\frac{K}{F_{s,t}}\right) = \ell_F\left(t, \log\frac{K}{\mathcal{X}_s} + \log\frac{\mathcal{X}_s}{F_{s,t}}\right) = \ell_F\left(t, \Bbbk + \log\frac{\mathcal{X}_s}{F_{s,t}}\right) \quad (158)$$

5.3.4. Equivalency in Formulas

The $\ell_{t,K}^2$ in (150), the $\ell_{t,\Bbbk}^2$ in (151) and the $\ell_{t,k}^2$ in (157) are in fact mutually equivalent. They are all stemmed from the classic Dupire local volatility expression (132). We show this by deriving its numerator and denominator in respective coordinate systems. For example, in (t, K, ξ) with Black-Scholes call $B = B_{t,K,\xi}$, the numerator and the denominator can be obtained using (149) and then (146) as

$$N = \frac{\partial \mathcal{C}_{t,K}}{\partial t} + \mu_t K \frac{\partial \mathcal{C}_{t,K}}{\partial K} + \hat{r}_t \mathcal{C}_{t,K} = \frac{\partial B}{\partial t} + \frac{\partial B}{\partial \xi}\frac{\partial \xi}{\partial t} + \mu_t K \frac{\partial B}{\partial K} + \mu_t K \frac{\partial B}{\partial \xi}\frac{\partial \xi}{\partial K} + \hat{r}_t B$$
$$= \frac{1}{2\xi\tau}\frac{\partial B}{\partial \xi}\left(\xi^2 + 2\xi\tau\left(\frac{\partial \xi}{\partial t} + \mu_t K \frac{\partial \xi}{\partial K}\right)\right) \quad (159)$$

$$D=\frac{1}{2}K^2\frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2}=\frac{1}{2}K^2\left(\frac{\partial^2 B}{\partial K^2}+2\frac{\partial^2 B}{\partial K\partial\xi}\frac{\partial\xi}{\partial K}+\frac{\partial B}{\partial\xi}\frac{\partial^2\xi}{\partial K^2}+\frac{\partial^2 B}{\partial\xi^2}\left(\frac{\partial\xi}{\partial K}\right)^2\right)$$

$$=\frac{1}{2\xi\tau}\frac{\partial B}{\partial\xi}\left(1+2\sqrt{\tau}Kd_+\frac{\partial\xi}{\partial K}+d_+d_-\tau K^2\left(\frac{\partial\xi}{\partial K}\right)^2+\xi\tau K^2\frac{\partial^2\xi}{\partial K^2}\right)$$

$$\frac{\partial B_{t,K,\xi}}{\partial\xi}=P_{s,t}F_{s,t}\phi(d_+)\sqrt{\tau}=P_{s,t}K\phi(d_-)\sqrt{\tau}$$

In $(t,\mathcal{k},\xi)$ with Black-Scholes call $B=B_{t,\mathcal{k},\xi}$, the numerator and denominator in (159) can be transformed equivalently using (152) into

$$N=\frac{1}{2\xi\tau}\frac{\partial B}{\partial\xi}\left(\xi^2+2\xi\tau\left(\frac{\partial\xi}{\partial t}+\mu_t\frac{\partial\xi}{\partial\mathcal{k}}\right)\right)$$

$$D=\frac{1}{2\xi\tau}\frac{\partial B}{\partial\xi}\left(\left(1-\frac{k}{\xi}\frac{\partial\xi}{\partial\mathcal{k}}\right)^2-\left(\frac{\xi\tau}{2}\frac{\partial\xi}{\partial\mathcal{k}}\right)^2+\xi\tau\frac{\partial^2\xi}{\partial\mathcal{k}^2}\right) \tag{160}$$

$$\frac{\partial B_{t,\mathcal{k},\xi}}{\partial\xi}=P_{s,t}F_{s,t}\phi(d_+)\sqrt{\tau}=P_{s,t}e^{\mathcal{k}}\mathcal{X}_s\phi(d_-)\sqrt{\tau}$$

Lastly in (t,k,v) with Black-Scholes call $B=B_{t,k,v}$, the respective numerator and denominator can be derived from (137), (156) and (154) as

$$N=\frac{\partial\mathcal{C}_{t,k}}{\partial t}+\hat{r}_t\mathcal{C}_{t,k}=\frac{\partial B}{\partial v}\frac{\partial v}{\partial t}$$

$$\begin{aligned}D&=\frac{1}{2}\left(\frac{\partial^2\mathcal{C}_{t,k}}{\partial k^2}-\frac{\partial\mathcal{C}_{t,k}}{\partial k}\right)=\frac{1}{2}\left(\frac{\partial^2 B}{\partial k^2}+2\frac{\partial^2 B}{\partial k\partial v}\frac{\partial v}{\partial k}+\frac{\partial B}{\partial v}\frac{\partial^2 v}{\partial k^2}+\frac{\partial^2 B}{\partial v^2}\left(\frac{\partial v}{\partial k}\right)^2-\frac{\partial B}{\partial k}-\frac{\partial B}{\partial v}\frac{\partial v}{\partial k}\right)\\&=\frac{\partial B}{\partial v}\left(1-\frac{k}{v}\frac{\partial v}{\partial k}+\frac{1}{4}\left(\frac{k^2}{v^2}-\frac{1}{v}-\frac{1}{4}\right)\left(\frac{\partial v}{\partial k}\right)^2+\frac{1}{2}\frac{\partial^2 v}{\partial k^2}\right)\end{aligned} \tag{161}$$

$$\frac{\partial B_{t,k,v}}{\partial v}=\frac{P_{s,t}F_{s,t}\phi(d_+)}{2\sqrt{v}}=\frac{P_{s,t}K\phi(d_-)}{2\sqrt{v}}$$

Alternatively, we may prove their equivalency directly. For example, this can be done as follows for the $\ell_{t,K}^2$ in (150) and the $\ell_{t,k}^2$ in (157)

$$
\begin{aligned}
\ell_{t,k}^2 &= \frac{\frac{\partial v}{\partial t}}{1-\frac{k}{v}\frac{\partial v}{\partial k}+\left(\frac{k^2}{4v^2}-\frac{1}{4v}-\frac{1}{16}\right)\left(\frac{\partial v}{\partial k}\right)^2+\frac{1}{2}\frac{\partial^2 v}{\partial k^2}} \\
&= \frac{\xi^2+2\xi\tau\left(\frac{\partial \xi}{\partial t}+\mu K\frac{\partial \xi}{\partial K}\right)}{1-2\xi\tau K\frac{k}{v}\frac{\partial \xi}{\partial K}+\left(\frac{k^2}{v^2}-\frac{1}{v}-\frac{1}{4}\right)\left(\xi\tau K\frac{\partial \xi}{\partial K}\right)^2+\tau K^2\left(\left(\frac{\partial \xi}{\partial K}\right)^2+\frac{\xi}{K}\frac{\partial \xi}{\partial K}+\xi\frac{\partial^2 \xi}{\partial K^2}\right)} \\
&= \frac{\xi^2+2\xi\tau\left(\frac{\partial \xi}{\partial t}+\mu K\frac{\partial \xi}{\partial K}\right)}{1+\left(1-2\frac{k}{v}\right)\xi\tau K\frac{\partial \xi}{\partial K}+\left(\frac{k^2}{v}-\frac{v}{4}\right)\tau K^2\left(\frac{\partial \xi}{\partial K}\right)^2+\xi\tau K^2\frac{\partial^2 \xi}{\partial K^2}} \\
&= \frac{\xi^2+2\xi\tau\left(\frac{\partial \xi}{\partial t}+\mu K\frac{\partial \xi}{\partial K}\right)}{1+2\sqrt{\tau}Kd_+\frac{\partial \xi}{\partial K}+d_+d_-\tau K^2\left(\frac{\partial \xi}{\partial K}\right)^2+\xi\tau K^2\frac{\partial^2 \xi}{\partial K^2}} = \ell_{t,K}^2
\end{aligned}
\tag{162}
$$

where by definition we have

$$
k=\log\frac{K}{F_{s,t}}, \qquad v=\xi^2\tau, \qquad d_\pm=\frac{\log\frac{F_{s,t}}{K}\pm\frac{\xi^2\tau}{2}}{\xi\sqrt{\tau}}=\frac{-k}{\sqrt{v}}\pm\frac{\sqrt{v}}{2}, \qquad d_+d_-=\frac{k^2}{v}-\frac{v}{4}
\tag{163}
$$

and also the following identities

$$
\begin{aligned}
\frac{\partial v}{\partial t} &= \frac{\partial(\xi^2\tau)}{\partial t}=\xi^2+2\xi\tau\left(\frac{\partial \xi}{\partial t}+\frac{\partial \xi}{\partial K}\frac{\partial K}{\partial t}\right)=\xi^2+2\xi\tau\left(\frac{\partial \xi}{\partial t}+\mu K\frac{\partial \xi}{\partial K}\right) \\
\frac{\partial v}{\partial k} &= \frac{\partial(\xi^2\tau)}{\partial k}=\frac{\partial(\xi^2\tau)}{\partial t}\frac{\partial t}{\partial k}+\frac{\partial(\xi^2\tau)}{\partial K}\frac{\partial K}{\partial k}=2\xi\tau\frac{\partial \xi}{\partial K}\frac{\partial K}{\partial k}=2\xi\tau K\frac{\partial \xi}{\partial K} \\
\frac{\partial^2 v}{\partial k^2} &= \frac{\partial\left(2\xi\tau K\frac{\partial \xi}{\partial K}\right)}{\partial t}\frac{\partial t}{\partial k}+\frac{\partial\left(2\xi\tau K\frac{\partial \xi}{\partial K}\right)}{\partial K}\frac{\partial K}{\partial k}=2\tau K\left(\xi\frac{\partial \xi}{\partial K}+K\frac{\partial \xi}{\partial K}\frac{\partial \xi}{\partial K}+\xi K\frac{\partial^2 \xi}{\partial K^2}\right) \\
&= 2\tau K^2\left(\left(\frac{\partial \xi}{\partial K}\right)^2+\frac{\xi}{K}\frac{\partial \xi}{\partial K}+\xi\frac{\partial^2 \xi}{\partial K^2}\right)
\end{aligned}
\tag{164}
$$

Notice that in (t,k)-plane the t and K are no longer independent and end up with the derivatives below

$$
\frac{\partial K}{\partial t}=\frac{\partial\left(F_{s,t}\exp(k)\right)}{\partial t}=\mu_t K, \qquad \frac{\partial K}{\partial k}=\frac{\partial\left(F_{s,t}\exp(k)\right)}{\partial k}=K
\tag{165}
$$

5.4. Forward Smile in Local Volatility

It is commonly observed that the implied volatility smile tends to flatten out as maturities become large. This observation can be explained in short as follows. Assuming that the FX spot is driven by a Brownian associated with a stochastic and mean reverting instantaneous volatility. Implied volatility can be thought of as an expectation of the time average of the instantaneous volatility. The instantaneous volatility evolves by nature in time wobbling around the mean reversion level to form a realized path. In other words, this path is composed of many samples of the mean reverting instantaneous volatility. If two samples of the instantaneous volatility are taken with enough time in between, they will appear independent of one another given that the mean reverting effect is sufficiently strong. When the maturity increases, it is as if we have more independent samples contributing to the average and therefore, we are likely to get a result closer to the true mean reversion level with a diminishing uncertainty. Mathematically speaking, the lower variance the average has, the flatter the smile will be [26].

It is ideal to use a simple model to further illustrate this finding. We may define a stochastic mean reverting process Y_t (known as *Ornstein-Uhlenbeck* process)

$$dY_t = -\kappa Y_t dt + \zeta dW_t, \qquad Y_s = 0 \tag{166}$$

where s is the initial time, $\kappa > 0$ the mean reverting rate, $\zeta > 0$ the volatility and W_t the driving Brownian. Since the Y_t process can be negative, it is not a perfect example of the instantaneous volatility, but still can be used as a good analogy for illustrative purpose. The analytical solution for Y_t is an Ito integral

$$Y_t = \zeta \int_s^t e^{-\kappa(t-v)} dW_v \tag{167}$$

With $\tau = t - s$, its time average can be derived as

$$\begin{aligned}\bar{Y}_t = \frac{1}{\tau}\int_s^t Y_u du = \frac{\zeta}{\tau}\int_s^t \int_s^u e^{-\kappa(u-v)} dW_v \, du = \frac{\zeta}{\tau}\int_s^t \int_v^t e^{-\kappa(u-v)} du \, dW_v \\ = \frac{\zeta}{\tau}\int_s^t \frac{1 - e^{-\kappa(t-v)}}{\kappa} dW_v\end{aligned} \tag{168}$$

The variance of the $\bar{Y}_t$ and its derivative can be computed as

$$
\begin{aligned}
\mathbb{V}[\bar{Y}_t] &= \frac{\zeta^2}{\tau^2}\int_s^t \left(\frac{1-e^{-\kappa(t-v)}}{\kappa}\right)^2 dv = \frac{\zeta^2}{\kappa^2\tau^2}\int_s^t \left(1-2e^{-\kappa(t-v)}+e^{-2\kappa(t-v)}\right)dv \\
&= \frac{\zeta^2}{\kappa^2\tau^2}\left(\tau - 2\frac{1-e^{-\kappa\tau}}{\kappa} + \frac{1-e^{-2\kappa\tau}}{2\kappa}\right) = \zeta^2\cdot\frac{2\kappa\tau-3+4e^{-\kappa\tau}-e^{-2\kappa\tau}}{2\kappa^3\tau^2} \\
\frac{\partial\mathbb{V}[\bar{Y}_t]}{\partial t} &= -\frac{\zeta^2}{\kappa^3\tau^3}\left(2\kappa\tau-3+4e^{-\kappa\tau}-e^{-2\kappa\tau}\right) + \frac{\zeta^2}{2\kappa^3\tau^2}\left(2\kappa-4\kappa e^{-\kappa\tau}+2\kappa e^{-2\kappa\tau}\right) \\
&= \zeta^2\cdot\left(\frac{e^{-2\kappa\tau}-4e^{-\kappa\tau}+3}{\kappa^3\tau^3} + \frac{e^{-2\kappa\tau}-2e^{-\kappa\tau}-1}{\kappa^2\tau^2}\right)
\end{aligned}
\tag{169}
$$

When $\kappa > 0$, we can see that (by L'Hôpital's rule)

$$
\lim_{t\to s}\mathbb{V}[\bar{Y}_t] = 0, \qquad \lim_{t\to s}\frac{\partial\mathbb{V}[\bar{Y}_t]}{\partial t} = \frac{\zeta^2}{3}, \qquad \lim_{t\to\infty}\mathbb{V}[\bar{Y}_t] = 0, \qquad \lim_{t\to\infty}\frac{\partial\mathbb{V}[\bar{Y}_t]}{\partial t} = 0 \tag{170}
$$

The above limits show that at very short maturities, the variance is close to zero and hence the smile is flat. This can be understood as the instantaneous volatility not having enough time to move much. As the maturities grows, the variance increases until reaching a maximum, and then decreases steadily. At sufficiently long maturities, the variance again goes to zero, resulting in a flat smile. Note that it is necessary to have non-zero mean reversion in the instantaneous volatility, so that the covariance $\mathbb{V}[Y_u, Y_v]$ between Y_u and Y_v for $s < u < v$ diminishes when v is sufficiently larger than u (in other words, the Y_u and Y_v become independent given enough time in between). This is not the case for $\kappa = 0$, where the variance will always grow, linearly in time, as shown below

$$
\lim_{\kappa\to 0}\bar{Y}_t = \frac{\zeta}{\tau}\int_s^t (t-v)dW_v\,, \qquad \lim_{\kappa\to 0}\mathbb{V}[\bar{Y}_t] = \frac{\zeta^2}{\tau^2}\int_s^t (t-v)^2 dv = \frac{\zeta^2\tau}{3} \tag{171}
$$

Skew of local volatility also tends to flatten out as maturity increases [27]. Using local volatility expression in (157), we see that the first derivative of local variance $\ell_{t,k}^2$ with respect to the strike equivalent quantity k can be derived as

$$
\frac{\partial\ell_{t,k}^2}{\partial k} = \frac{1}{D}\frac{\partial}{\partial k}\frac{\partial v}{\partial t} - \frac{1}{D^2}\frac{\partial v}{\partial t}\frac{\partial D}{\partial k} \tag{172}
$$

where v is the total implied variance and

$$D = 1 - \frac{k}{v}\frac{\partial v}{\partial k} + \frac{1}{4}\left(\frac{k^2}{v^2} - \frac{1}{v} - \frac{1}{4}\right)\left(\frac{\partial v}{\partial k}\right)^2 + \frac{1}{2}\frac{\partial^2 v}{\partial k^2}$$

$$\begin{aligned}\frac{\partial D}{\partial k} = &-\frac{1}{v}\frac{\partial v}{\partial k} + \frac{k}{v^2}\left(\frac{\partial v}{\partial k}\right)^2 - \frac{k}{v}\frac{\partial^2 v}{\partial k^2} + \frac{k}{2v^2}\left(\frac{\partial v}{\partial k}\right)^2 - \frac{k^2}{2v^3}\left(\frac{\partial v}{\partial k}\right)^3 + \frac{1}{4v^2}\left(\frac{\partial v}{\partial k}\right)^3 \\ &+ \frac{1}{2}\left(\frac{k^2}{v^2} - \frac{1}{v} - \frac{1}{4}\right)\frac{\partial v}{\partial k}\frac{\partial^2 v}{\partial k^2} + \frac{1}{2}\frac{\partial^3 v}{\partial k^3}\end{aligned} \tag{173}$$

From the analysis in previous paragraph, we know that at long maturities (i.e., t is sufficiently large), the quantity $\partial v/\partial k$ becomes small, and the following quantities approach to zero at even faster rates

$$\left(\frac{\partial v}{\partial k}\right)^2 \to 0, \qquad \left(\frac{\partial v}{\partial k}\right)^3 \to 0, \qquad \frac{\partial^2 v}{\partial k^2} \to 0, \qquad \frac{\partial^3 v}{\partial k^3} \to 0 \tag{174}$$

This allows us to approximate (173) by dropping these negligible terms

$$D \approx 1, \qquad \frac{\partial D}{\partial k} \approx -\frac{1}{v}\frac{\partial v}{\partial k} \tag{175}$$

and further approximate (172) by

$$\frac{\partial \ell_{t,k}^2}{\partial k} \approx \left(\frac{\partial}{\partial t} + \frac{1}{v}\frac{\partial v}{\partial t}\right)\frac{\partial v}{\partial k} \tag{176}$$

The (176) tells that the skew of local volatility depends on the skew of implied volatility. As implied volatility flattens out at long maturities, so does the local volatility, leading to a flattening of the forward smile (i.e., the smile in the future), which is unrealistic. This is not desirable when an exotic option has considerable exposure to the forward smile.

6. Local Volatility with Stochastic Rates: General Dupire

A common extension of the classic local volatility model is to include stochastic rates in the model dynamics, resulting in stochastic drift term of the FX spot. Terminal distribution of the spot has now dependence not only on the diffusion term characterized by the spot volatility, but also on the stochastic rates through the drift term. We want to find a general local volatility that is able to reproduce the terminal distribution while taking the stochastic rates into account.

6.1. General Dupire Local Volatility

Suppose that the domestic short rate r_t and the foreign short rate $\hat{r}_t$ are stochastic, the FX spot process follows a general SDE

$$\frac{d\mathcal{X}_t}{\mathcal{X}_t} = (r_t - \hat{r}_t)dt + \sigma_t dW_t \tag{177}$$

where the volatility process σ_t can be stochastic and may also be dependent on spot $\mathcal{X}_t$. We can derive the following identities for the call and put option

$$\begin{aligned}
\mathcal{C}_{t,K} &= \mathbb{E}_s\left[D_{s,t}(\mathcal{X}_t - K)^+\right] = \mathbb{E}_s\left[D_{s,t}(\mathcal{X}_t - K)\Theta(\mathcal{X}_t - K)\right] \\
&= \mathbb{E}_s\left[D_{s,t}\mathcal{X}_t\Theta(\mathcal{X}_t - K)\right] + K\frac{\partial \mathcal{C}_{t,K}}{\partial K} \\
\frac{\partial \mathcal{C}_{t,K}}{\partial K} &= -\mathbb{E}_s\left[D_{s,t}\Theta(\mathcal{X}_t - K)\right] \\
\frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2} &= \mathbb{E}_s\left[D_{s,t}\delta(\mathcal{X}_t - K)\right] = \mathbb{E}_s\left[D_{s,t}\middle|\mathcal{X}_t = K\right]\mathbb{E}_s[\delta(\mathcal{X}_t - K)] \\
\mathcal{P}_{t,K} &= \mathbb{E}_s\left[D_{s,t}(K - \mathcal{X}_t)^+\right] = \mathbb{E}_s\left[D_{s,t}(K - \mathcal{X}_t)\Theta(K - \mathcal{X}_t)\right] \\
&= K\frac{\partial \mathcal{P}_{t,K}}{\partial K} - \mathbb{E}_s\left[D_{s,t}\mathcal{X}_t\Theta(K - \mathcal{X}_t)\right] \\
\frac{\partial \mathcal{P}_{t,K}}{\partial K} &= \mathbb{E}_s\left[D_{s,t}\Theta(K - \mathcal{X}_t)\right] \\
\frac{\partial^2 \mathcal{P}_{t,K}}{\partial K^2} &= \mathbb{E}_s\left[D_{s,t}\delta(K - \mathcal{X}_t)\right] = \mathbb{E}_s\left[D_{s,t}\middle|\mathcal{X}_t = K\right]\mathbb{E}_s[\delta(K - \mathcal{X}_t)]
\end{aligned} \tag{178}$$

where Θ is the *Heaviside* step function and δ is the *Dirac* delta function. Further using Ito-Tanaka formula (25) on the (non-smooth) terminal payoff function, we get

$$\begin{aligned}
d\left(D_{s,t}(\mathcal{X}_t - K)^+\right) &= -r_t D_{s,t}(\mathcal{X}_t - K)^+ dt + D_{s,t}\Theta(\mathcal{X}_t - K)d\mathcal{X}_t + \frac{D_{s,t}\ \delta(\mathcal{X}_t - K)d\langle\mathcal{X},\mathcal{X}\rangle_t}{2} \\
d\left(D_{s,t}(K - \mathcal{X}_t)^+\right) &= -r_t D_{s,t}(K - \mathcal{X}_t)^+ dt - D_{s,t}\Theta(K - \mathcal{X}_t)d\mathcal{X}_t + \frac{D_{s,t}\ \delta(K - \mathcal{X}_t)d\langle\mathcal{X},\mathcal{X}\rangle_t}{2}
\end{aligned} \tag{179}$$

By Fubini's theorem, this allows us to write for the call option

$$
\begin{aligned}
\frac{\partial \mathcal{C}_{t,K}}{\partial t} &= \mathbb{E}_s\left[\frac{d\left(D_{s,t}(\mathcal{X}_t-K)^+\right)}{dt}\right] \\
&= \mathbb{E}_s\left[D_{s,t}\left(-r_t(\mathcal{X}_t-K)\Theta(\mathcal{X}_t-K)+(r_t-\hat{r}_t)\mathcal{X}_t\Theta(\mathcal{X}_t-K)+\frac{1}{2}\mathcal{X}_t^2\sigma_t^2\delta(\mathcal{X}_t-K)\right)\right] \\
&= -\mathbb{E}_s\left[D_{s,t}(\hat{r}_t\mathcal{X}_t-r_tK)\Theta(\mathcal{X}_t-K)\right]+\frac{1}{2}\mathbb{E}_s\left[D_{s,t}\mathcal{X}_t^2\sigma_t^2\delta(\mathcal{X}_t-K)\right] \\
&= -\mathbb{E}_s\left[D_{s,t}(\hat{r}_t\mathcal{X}_t-r_tK)\Theta(\mathcal{X}_t-K)\right]+\frac{1}{2}K^2\mathbb{E}_s\left[D_{s,t}\sigma_t^2\middle|\mathcal{X}_t=K\right]\mathbb{E}_s[\delta(\mathcal{X}_t-K)] \\
&= -\mathbb{E}_s\left[D_{s,t}(\hat{r}_t\mathcal{X}_t-r_tK)\Theta(\mathcal{X}_t-K)\right]+\frac{1}{2}K^2\frac{\partial^2\mathcal{C}_{t,K}}{\partial K^2}\frac{\mathbb{E}_s\left[D_{s,t}\sigma_t^2\middle|\mathcal{X}_t=K\right]}{\mathbb{E}_s\left[D_{s,t}\middle|\mathcal{X}_t=K\right]}
\end{aligned} \tag{180}
$$

and for the put option

$$
\begin{aligned}
\frac{\partial \mathcal{P}_{t,K}}{\partial t} &= \mathbb{E}_s\left[\frac{d\left(D_{s,t}(K-\mathcal{X}_t)^+\right)}{dt}\right] \\
&= \mathbb{E}_s\left[D_{s,t}\left(-r_t(K-\mathcal{X}_t)\Theta(K-\mathcal{X}_t)-(r_t-\hat{r}_t)\mathcal{X}_t\Theta(K-\mathcal{X}_t)+\frac{1}{2}\mathcal{X}_t^2\sigma_t^2\delta(K-\mathcal{X}_t)\right)\right] \\
&= -\mathbb{E}_s\left[D_{s,t}(r_tK-\hat{r}_t\mathcal{X}_t)\Theta(K-\mathcal{X}_t)\right]+\frac{1}{2}\mathbb{E}_s\left[D_{s,t}\mathcal{X}_t^2\sigma_t^2\delta(K-\mathcal{X}_t)\right] \\
&= -\mathbb{E}_s\left[D_{s,t}(r_tK-\hat{r}_t\mathcal{X}_t)\Theta(K-\mathcal{X}_t)\right]+\frac{1}{2}K^2\mathbb{E}_s\left[D_{s,t}\sigma_t^2\middle|\mathcal{X}_t=K\right]\mathbb{E}_s[\delta(K-\mathcal{X}_t)] \\
&= -\mathbb{E}_s\left[D_{s,t}(r_tK-\hat{r}_t\mathcal{X}_t)\Theta(K-\mathcal{X}_t)\right]+\frac{1}{2}K^2\frac{\partial^2\mathcal{P}_{t,K}}{\partial K^2}\frac{\mathbb{E}_s\left[D_{s,t}\sigma_t^2\middle|\mathcal{X}_t=K\right]}{\mathbb{E}_s\left[D_{s,t}\middle|\mathcal{X}_t=K\right]}
\end{aligned} \tag{181}
$$

where given the delta function, we have used the fact that

$$
\begin{aligned}
&\mathbb{E}_s\left[D_{s,t}\mathcal{X}_t^2\sigma_t^2\delta(\mathcal{X}_t-K)\right]=K^2\mathbb{E}_s\left[D_{s,t}\sigma_t^2\middle|\mathcal{X}_t=K\right]\mathbb{E}_s[\delta(\mathcal{X}_t-K)] \\
&\mathbb{E}_s\left[D_{s,t}\mathcal{X}_t^2\sigma_t^2\delta(\mathcal{X}_t-K)\right]=\int_\Omega D_{s,t}\mathcal{X}_t^2\sigma_t^2\delta(\mathcal{X}_t-K)p_\Omega(\omega)d\omega \\
&= \int_\mathcal{X}\int_\Pi D_{s,t}\mathcal{X}_t^2\sigma_t^2\delta(\mathcal{X}_t-K)p_\Pi(\pi|\mathcal{X}=x)p_\mathcal{X}(\mathcal{X}=x)d\pi\, dx \\
&= \left(\int_\Pi D_{s,t}K^2\sigma_t^2p_\Pi(\pi|\mathcal{X}=K)d\pi\right)\left(\int_\mathcal{X}\delta(\mathcal{X}_t-K)p_\mathcal{X}(\mathcal{X}=x)dx\right)
\end{aligned} \tag{182}
$$

$$= K^2 \mathbb{E}_s\left[D_{s,t}\sigma_t^2 \middle| \mathcal{X}_t = K\right] \mathbb{E}_s[\delta(\mathcal{X}_t - K)]$$

Basically the (180) and (181) tell that

$$\begin{aligned} \frac{\mathbb{E}_s\left[D_{s,t}\sigma_t^2 \middle| \mathcal{X}_t = K\right]}{\mathbb{E}_s\left[D_{s,t} \middle| \mathcal{X}_t = K\right]} &= \frac{\frac{\partial \mathcal{C}_{t,K}}{\partial t} + \mathbb{E}_s\left[D_{s,t}(\hat{r}_t \mathcal{X}_t - r_t K)\Theta(\mathcal{X}_t - K)\right]}{\frac{1}{2}K^2 \frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2}} \\ &= \frac{\frac{\partial \mathcal{P}_{t,K}}{\partial t} + \mathbb{E}_s\left[D_{s,t}(r_t K - \hat{r}_t \mathcal{X}_t)\Theta(K - \mathcal{X}_t)\right]}{\frac{1}{2}K^2 \frac{\partial^2 \mathcal{P}_{t,K}}{\partial K^2}} \end{aligned} \tag{183}$$

If we write the volatility term σ_t as a pure local volatility $\mathscr{g}(t, \mathcal{X}_t)$, which is a deterministic function of $\mathcal{X}_t$, the σ_t^2 can then be moved out of the expectation and the discount factor cancels. Hence, under the condition of stochastic rates, we obtain

$$\mathscr{g}_{t,K}^2 = \frac{\frac{\partial \mathcal{C}_{t,K}}{\partial t} + \mathbb{E}_s\left[D_{s,t}(\hat{r}_t \mathcal{X}_t - r_t K)\Theta(\mathcal{X}_t - K)\right]}{\frac{1}{2}K^2 \frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2}} = \frac{\frac{\partial \mathcal{P}_{t,K}}{\partial t} + \mathbb{E}_s\left[D_{s,t}(r_t K - \hat{r}_t \mathcal{X}_t)\Theta(K - \mathcal{X}_t)\right]}{\frac{1}{2}K^2 \frac{\partial^2 \mathcal{P}_{t,K}}{\partial K^2}} \tag{184}$$

This is the *general* Dupire local volatility, expressed in call or put option, respectively. Please refer to [28] for an introduction of the topic.

If we again assume the short rates $r_t = f_{s,t}$ and $\hat{r}_t = \hat{f}_{s,t}$ are deterministic (i.e., they are the instantaneous forward rates as in classic Dupire local volatility), the (183) simplifies (using relations in (178)) to the classic Dupire local volatility formula (132) and (133), that is

$$\begin{aligned} \ell_{t,K}^2 &= \mathbb{E}_s[\sigma_t^2 | \mathcal{X}_t = K] \\ &= \frac{\frac{\partial \mathcal{C}_{t,K}}{\partial t} + \left(f_{s,t} - \hat{f}_{s,t}\right)K \frac{\partial \mathcal{C}_{t,K}}{\partial K} + \hat{f}_{s,t}\mathcal{C}_{t,K}}{\frac{1}{2}K^2 \frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2}} = \frac{\frac{\partial \mathcal{P}_{t,K}}{\partial t} + \left(f_{s,t} - \hat{f}_{s,t}\right)K \frac{\partial \mathcal{P}_{t,K}}{\partial K} + \hat{f}_{s,t}\mathcal{P}_{t,K}}{\frac{1}{2}K^2 \frac{\partial^2 \mathcal{P}_{t,K}}{\partial K^2}} \end{aligned} \tag{185}$$

This indicates that the conditional expectation of the instantaneous stochastic variance is equal to the classic Dupire local variance [29]. In other words, local variance is the risk-neutral expectation of the instantaneous variance conditional on the final spot $\mathcal{X}_t$ equal to K [30].

By further defining (centered) rates as

$$\lambda_t = r_t - f_{s,t}, \qquad \hat{\lambda}_t = \hat{r}_t - \hat{f}_{s,t} \tag{186}$$

we can write the difference between the general Dupire local volatility $g_{t,K}$ (184) and the classic Dupire local volatility $\ell_{t,K}$ (132) and (133) as

$$\gamma_{t,K} = g_{t,K}^2 - \ell_{t,K}^2 = \frac{\mathbb{E}_s\left[D_{s,t}\left(\hat{\lambda}_t\mathcal{X}_t - \lambda_t K\right)\Theta(\mathcal{X}_t - K)\right]}{\frac{1}{2}K^2\frac{\partial^2\mathcal{C}_{t,K}}{\partial K^2}} = \frac{\mathbb{E}_s\left[D_{s,t}\left(\lambda_t K - \hat{\lambda}_t\mathcal{X}_t\right)\Theta(K - \mathcal{X}_t)\right]}{\frac{1}{2}K^2\frac{\partial^2\mathcal{P}_{t,K}}{\partial K^2}} \tag{187}$$

As one can see, the difference $\gamma_{t,K}$ is determined by the 3D joint distribution of FX spot and both rates.

Since that the two denominators in (187) are equal (from (178) and (125))

$$\frac{\partial^2\mathcal{C}_{t,K}}{\partial K^2} = \frac{\partial^2\mathcal{P}_{t,K}}{\partial K^2} = \mathbb{E}_s\left[D_{s,t}\delta(\mathcal{X}_t - K)\right] = P_{s,t}p_{t,K} \quad \text{and} \quad \Theta(K - \mathcal{X}_t) = 1 - \Theta(\mathcal{X}_t - K) \tag{188}$$

the two numerators must also be equal. Given the fact that

$$\Theta(K - \mathcal{X}_t) = 1 - \Theta(\mathcal{X}_t - K) \tag{189}$$

we can show that

$$\begin{aligned}
&\mathbb{E}_s\left[D_{s,t}\left(\hat{\lambda}_t\mathcal{X}_t - \lambda_t K\right)\Theta(\mathcal{X}_t - K)\right] = \mathbb{E}_s\left[D_{s,t}\left(\lambda_t K - \hat{\lambda}_t\mathcal{X}_t\right)\Theta(K - \mathcal{X}_t)\right] \\
&= -\mathbb{E}_s\left[D_{s,t}\left(\hat{\lambda}_t\mathcal{X}_t - \lambda_t K\right)\left(1 - \Theta(\mathcal{X}_t - K)\right)\right] \\
&= \mathbb{E}_s\left[D_{s,t}\left(\hat{\lambda}_t\mathcal{X}_t - \lambda_t K\right)\Theta(\mathcal{X}_t - K)\right] - \mathbb{E}_s\left[D_{s,t}\left(\hat{\lambda}_t\mathcal{X}_t - \lambda_t K\right)\right] \\
&\Longrightarrow \quad \mathbb{E}_s\left[D_{s,t}\left(\hat{\lambda}_t\mathcal{X}_t - \lambda_t K\right)\right] = \mathbb{E}_s\left[D_{s,t}\hat{\lambda}_t\mathcal{X}_t\right] - K\mathbb{E}_s\left[D_{s,t}\lambda_t\right] = 0
\end{aligned} \tag{190}$$

The (190) can also be shown by simply observing that

$$\begin{aligned}
&\mathbb{E}_s\left[D_{s,t}\lambda_t\right] = \mathbb{E}_s\left[D_{s,t}r_t\right] - f_{s,t}\mathbb{E}_s\left[D_{s,t}\right] = 0 \quad \text{where} \\
&\mathbb{E}_s\left[D_{s,t}\right] = P_{s,t}, \qquad \mathbb{E}_s\left[D_{s,t}r_t\right] = -\mathbb{E}_s\left[\frac{dD_{s,t}}{dt}\right] = -\frac{d}{dt}\mathbb{E}_s\left[D_{s,t}\right] = -\frac{dP_{s,t}}{dt} = f_{s,t}P_{s,t}
\end{aligned} \tag{191}$$

and

$$\begin{aligned}
&\mathbb{E}_s\left[D_{s,t}\hat{\lambda}_t\mathcal{X}_t\right] = \mathbb{E}_s\left[D_{s,t}\hat{r}_t\mathcal{X}_t\right] - \hat{f}_{s,t}\mathbb{E}_s\left[D_{s,t}\mathcal{X}_t\right] = 0 \quad \text{where} \\
&\mathbb{E}_s\left[D_{s,t}\hat{r}_t\mathcal{X}_t\right] = \mathcal{X}_s\widehat{\mathbb{E}}_s\left[\widehat{D}_{s,t}\hat{r}_t\right] = \mathcal{X}_s\hat{f}_{s,t}\widehat{P}_{s,t} = \hat{f}_{s,t}P_{s,t}F_{s,t}
\end{aligned} \tag{192}$$

$$\mathbb{E}_s\left[D_{s,t}\mathcal{X}_t\right] = M_s\mathbb{E}_s\left[\frac{\hat{P}_{t,t}\mathcal{X}_t}{M_t}\right] = P_{s,t}\mathbb{E}_s^t\left[\frac{\hat{P}_{t,t}\mathcal{X}_t}{P_{t,t}}\right] = P_{s,t}\frac{\hat{P}_{s,t}\mathcal{X}_s}{P_{s,t}} = P_{s,t}F_{s,t}$$

From (188) and (190) we can see that both numerator and denominator of (187) vanish at wings of volatility smile (i.e., K is either very low or very high), this will demands special treatment to avoid divided-by-zero when estimating the $\gamma_{t,K}$ numerically.

6.2. The General Dupire Model

We may construct a 3-factor model that features a local volatility (e.g., the general Dupire local volatility $\mathcal{g}_{t,X}$) for FX spot and 1 factor Hull-White dynamics for rates

$$\begin{gathered}
r_t = f_{s,t} + \int_s^t b_{u,t}\beta_{u,t}du + x_t, \qquad x_t = \int_s^t \beta_{u,t}dW_{2,u}\,, \qquad dx_t = -\mu x_t dt + \varsigma_t dW_{2,t} \\
\hat{r}_t = \hat{f}_{s,t} + \int_s^t \hat{b}_{u,t}\hat{\beta}_{u,t}du + \hat{x}_t, \qquad \hat{x}_t = -\int_s^t \rho_u^{01}\hat{\beta}_{u,t}\mathcal{g}_{u,X}du + \int_s^t \hat{\beta}_{u,t}dW_{1,u}\,, \\
d\hat{x}_t = -\left(\rho_t^{01}\hat{\varsigma}_t\mathcal{g}_{t,X} + \hat{\mu}\hat{x}_t\right)dt + \hat{\varsigma}_t dW_{1,t} \\
dX_t = \left(\lambda_t - \hat{\lambda}_t - \frac{1}{2}\mathcal{g}_{t,X}^2\right)dt + \mathcal{g}_{t,X}dW_{0,t}, \qquad dW_{i,t}dW_{j,t} = \rho_t^{ij}dt \;\;\forall\;\; i,j = 0,1,2
\end{gathered} \tag{193}$$

where X_t, $\hat{x}_t$ and x_t are the state variables. To minimize the drifting, we define $X_t \equiv \log(\mathcal{X}_t/F_{s,t})$ the log forward moneyness with FX spot $\mathcal{X}_t$ and forward $F_{s,t} = \mathcal{X}_s\hat{P}_{s,t}/P_{s,t}$. The (centered) rates λ_t and $\hat{\lambda}_t$ are given in (186), where r_t and $\hat{r}_t$ are the domestic and foreign short rates respectively (where a variable with a "hat" accent denotes a quantity associated with foreign economy).

Table 4. Notations in the general Dupire local volatility model

Notation	Description
$\mathcal{X}_t$	FX spot
$F_{s,t}$	FX forward, $F_{s,t} = \mathcal{X}_s\hat{P}_{s,t}/P_{s,t}$
$P_{s,t}$, $\hat{P}_{s,t}$	domestic and foreign zero-coupon bond
x_t, $\hat{x}_t$	domestic and foreign rates state variable
r_t, $\hat{r}_t$	domestic and foreign short rates
$f_{s,t}$, $\hat{f}_{s,t}$	instantaneous forward rates
λ_t, $\hat{\lambda}_t$	centered short rates, $\lambda_t = r_t - f_{s,t}$
$\beta_{u,t}$, $\hat{\beta}_{u,t}$	volatility of forward rates
$b_{u,t}$, $\hat{b}_{u,t}$	volatility of zero-coupon bond
μ, $\hat{\mu}$	constant mean reversion of short rates

ς_t, $\hat{\varsigma}_t$	volatility of short rates
X_t	FX log forward moneyness, $X_t = \log\left(\mathcal{X}_t / F_{s,t}\right)$
$\mathcal{g}(t, X_t)$	general Dupire local volatility
ρ_t^{01}, ρ_t^{02}, ρ_t^{12}	correlation between X_t, $\hat{r}_t$ and r_t

6.3. Stochastic Rates: One-Factor Hull White Model

In the Hull-White model[1], the short rate r_t and its driving process x_t are given by

$$r_t = f_{s,t} + \int_s^t b_{u,t}\beta_{u,t}du + x_t, \qquad x_t = \int_s^t \beta_{u,t}dW_u, \qquad dx_t = -\mu x_t dt + \varsigma_t dW_t \tag{194}$$

where $\beta_{u,t}$ is the volatility of the instantaneous forward rate $f_{u,t}$, and $b_{u,t}$ the volatility of the zero coupon bond $P_{u,t}$, respectively, which are given below

$$\begin{aligned} &\beta_{u,t} = e^{-\mu(t-u)}\varsigma_u, \qquad b_{u,t} = \int_u^t \beta_{u,v}dv = \frac{1-e^{-\mu(t-u)}}{\mu}\varsigma_u, \qquad b_{u,T} - b_{u,t} = b_{t,T}\beta_{u,t} \\ &\lim_{\mu\to 0}\beta_{u,t} = \varsigma_u, \qquad \lim_{\mu\to 0} b_{u,t} = (t-u)\varsigma_u \end{aligned} \tag{195}$$

In practical applications, the model (194) usually takes a time-invariant μ, which is often an exogenous model input, and calibrates (deterministic) piecewise constant term structure ς, such that $\varsigma_t = \varsigma_i \ \forall\ t_{i-1} < t \le t_i$, to caplets or co-terminal swaptions. The reason a time variant μ is *not* in favor is that it makes the evolution of forward rate volatility strongly non-stationary. This has been intensively discussed in [31]. Writing the model (194) in a more familiar form, it would look like

$$dr_t = \mu(\theta_t - r_t)dt + \varsigma_t dW_t, \qquad \theta_t = f_{s,t} + \frac{1}{\mu}\frac{\partial f_{s,t}}{\partial t} + \frac{1}{\mu}\int_s^t \beta_{u,t}^2 du \tag{196}$$

Below we derive the integrals for the variance and covariance, which are further expressed in summations providing the ς_u is piecewise constant in time

[1] Formal derivation of the model can be found in my notes "*Introduction to Interest Rate Models*", which can be obtained from https://modelmania.github.io/main/

$$\begin{aligned}
\chi_{s,t} = \int_s^t b_{u,t}\beta_{u,t}du = \int_s^t \varsigma_u^2 e^{-\mu(t-u)}\frac{1-e^{-\mu(t-u)}}{\mu}du \\
= \frac{e^{-\mu t}}{\mu}\int_s^t \varsigma_u^2 e^{\mu u}du - \frac{e^{-2\mu t}}{\mu}\int_s^t \varsigma_u^2 e^{2\mu u}du \\
= \frac{e^{-\mu t}}{\mu^2}\sum_{j=s}^t \varsigma_j^2\left(e^{\mu t_j} - e^{\mu t_{j-1}}\right) - \frac{e^{-2\mu t}}{2\mu^2}\sum_{j=s}^t \varsigma_j^2\left(e^{2\mu t_j} - e^{2\mu t_{j-1}}\right) \\
\lim_{\mu\to 0}\chi_{s,t} = \int_s^t (t-u)\varsigma_u^2 du = t\sum_{j=s}^t \varsigma_j^2\left(t_j - t_{j-1}\right) - \frac{1}{2}\sum_{j=s}^t \varsigma_j^2\left(t_j - t_{j-1}\right)^2 \\
\varphi_{s,t} = \int_s^t \beta_{u,t}^2 du = \int_s^t \varsigma_u^2 e^{-2\mu(t-u)}du = e^{-2\mu t}\int_s^t \varsigma_u^2 e^{2\mu u}du = \frac{e^{-2\mu t}}{2\mu}\sum_{j=s}^t \varsigma_j^2\left(e^{2\mu t_j} - e^{2\mu t_{j-1}}\right) \\
\lim_{\mu\to 0}\varphi_{s,t} = \sum_{j=s}^t \varsigma_j^2\left(t_j - t_{j-1}\right)
\end{aligned} \tag{197}$$

The time t zero coupon bond for a maturity T admits an expression as follows

$$\begin{aligned}
P_{t,T} = \frac{P_{s,T}}{P_{s,t}}\exp\left(-\int_s^t \frac{b_{u,T}^2 - b_{u,t}^2}{2}du - \int_s^t \left(b_{u,T} - b_{u,t}\right)dW_u\right) \\
= \frac{P_{s,T}}{P_{s,t}}\exp\left(-\int_s^t \frac{\left(b_{u,T} - b_{u,t}\right)^2 + 2b_{u,t}\left(b_{u,T} - b_{u,t}\right)}{2}du - b_{t,T}x_t\right) \\
= \frac{P_{s,T}}{P_{s,t}}\exp\left(-\frac{b_{t,T}^2}{2}\int_s^t \beta_{u,t}^2 du - b_{t,T}\int_s^t b_{u,t}\beta_{u,t}du - b_{t,T}x_t\right) \\
= \frac{P_{s,T}}{P_{s,t}}\exp\left(-\frac{1}{2}b_{t,T}^2\varphi_{s,t} - b_{t,T}\chi_{s,t} - b_{t,T}x_t\right)
\end{aligned} \tag{198}$$

Model calibration relies on the fact that a forward starting zero coupon bond under T-forward measure is a lognormal martingale, that is

$$
\begin{aligned}
P_{t,T,V} &= \frac{P_{t,V}}{P_{t,T}} = \frac{P_{s,V}}{P_{s,T}} \exp\left(-\int_s^t \frac{b_{u,V}^2 - b_{u,T}^2}{2} du - \int_s^t \left(b_{u,V} - b_{u,T}\right) dW_u\right) \\
&= \frac{P_{s,V}}{P_{s,T}} \exp\left(-\int_s^t \frac{\left(b_{u,V} - b_{u,T}\right)^2}{2} du - \int_s^t \left(b_{u,V} - b_{u,T}\right) dW_u^T\right)
\end{aligned}
\tag{199}
$$

where we have used the change of measure

$$
dW_u^T = dW_u + b_{u,T} du \tag{200}
$$

We may also write

$$
P_{t,T,V} = \frac{P_{t,V}}{P_{t,T}} = \frac{P_{s,V}}{P_{s,T}} \exp\left(-\frac{1}{2}\xi_{s,t,T,V}^2 - \xi_{s,t,T,V} Z_t\right) \tag{201}
$$

where Z_t is a standard normal random variable and the volatility $\xi_{s,t,T,V}$ is defined as follows

$$
\begin{aligned}
&\xi_{s,t,T,V}^2 = \int_s^t \left(b_{u,V} - b_{u,T}\right)^2 du = \int_s^t \left(\frac{e^{-\mu(T-u)} - e^{-\mu(V-u)}}{\mu}\right)^2 \varsigma_u^2 du \\
&= \left(\frac{e^{-\mu T} - e^{-\mu V}}{\mu}\right)^2 \int_s^t e^{2\mu u} \varsigma_u^2 du = \left(\frac{e^{-\mu T} - e^{-\mu V}}{\mu}\right)^2 \frac{1}{2\mu} \sum_{j=s}^t \varsigma_j^2 \left(e^{2\mu t_j} - e^{2\mu t_{j-1}}\right)
\end{aligned}
\tag{202}
$$

When $\mu \to 0$, we have the limiting case

$$
\lim_{\mu \to 0} \xi_{s,t,T,V}^2 = (V-T)^2 \sum_{j=s}^t \varsigma_j^2 \left(t_j - t_{j-1}\right) \tag{203}
$$

The calibration to caplets is trivial. Below we will focus on the calibration to co-terminal swaptions.

6.3.1. Model Value of Swaption: Jamshidian Decomposition

The payer swaption traded at s and expired at $t \le t_a$ can be priced by (after change of measure from $\mathbb{Q}$ to $\mathbb{Q}^t$)

$$
\begin{aligned}
&V_{s,t,a,b}^{\mathrm{PS}} = P_{s,t} \mathbb{E}_s^t \left[\left(\sum_{i=a+1}^b P_{t,i} \tau_i \left(L_{t,i} - K\right)\right)^+\right] = P_{s,t} \mathbb{E}_s^t \left[\left(P_{t,a} - P_{t,b} - K \sum_{i=a+1}^b P_{t,i} \tau_i\right)^+\right] \\
&= P_{s,t} \mathbb{E}_s^t \left[\left(\sum_{i=a}^b c_i P_{t,i}\right)^+\right], \qquad c_i = \begin{cases} 1 & \text{if } i = a \\ -\tau_i K & \text{if } a+1 \le i \le b-1 \\ -1 - \tau_i K & \text{if } i = b \end{cases}
\end{aligned}
\tag{204}
$$

with the forward bond price

$$P_{t,i} = \frac{P_{s,i}}{P_{s,t}} \exp\left(-\frac{1}{2}\xi_{s,t,t,i}^2 - \xi_{s,t,t,i} Z_t\right) \tag{205}$$

As can be seen, the forward coupon bonds $P_{t,i}$ with different maturity t_i are all driven by a common standard normal random variable Z_t. This affine term structure allows us to compute swaption value in the model using a method proposed by Henrard [32] in 2003, which is basically a variant of the Jamshidian's decomposition.

We elaborate the method below. In the one-factor Hull-White model, the (204) can be written as

$$V_{s,t,a,b}^{\mathrm{PS}} = \int_{\mathbb{R}} \left(\sum_{i=a}^{b} \delta_i \exp(-\xi_i z)\right)^+ \phi(z) dz\,, \qquad \delta_i = c_i P_{s,i} \exp\left(-\frac{1}{2}\xi_i^2\right) \tag{206}$$

where $\xi_i = \xi_{s,t,t,i}$ for brevity and $\phi(z)$ is the standard normal density function

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \tag{207}$$

Let $h(z)$ be the payer swap payoff function in (206), that is

$$h(z) = \sum_{i=a}^{b} \delta_i \exp(-\xi_i z) \tag{208}$$

The $h(z)$ can be regarded as a sum of exponentially decayed δ_i with non-negative decaying factor ξ_i. Since δ_i has the same sign of c_i, we can imagine that the δ_i's are all positive up to a certain $i = k$ (e.g., $i = a$ for positive K or $i = b - 1$ for negative K), then all negative. Let's define another axillary function $g(z)$

$$g(z) = h(z) \exp(\xi_k z) = \sum_{i=a}^{b} \delta_i \exp\big((\xi_k - \xi_i) z\big) \tag{209}$$

Because ξ_i is monotonically increasing as bond maturity grows (i.e., $\xi_i < \xi_{i+1}$ for $t_{i+1} > t_i$), the δ_i and $(\xi_k - \xi_i)$ now have the same sign. Therefore $g(z)$ is strictly increasing. Since $g(z)$ is negative when $z \to -\infty$ and positive when $z \to +\infty$, the monotonicity in $g(z)$ ensures that there is one and only one solution

z^* such that $g(z^*) = 0$, and so is it for $h(z)$. In other words, given the unique z^*, the $h(z) < 0$ if $z < z^*$ and $h(z) \geq 0$ otherwise. Hence the payer swaption price in (206) can be transformed into

$$\begin{aligned} V_{s,t,a,b}^{\mathrm{PS}} &= \int_{z^*}^{\infty} \sum_{i=a}^{b} \delta_i \exp(-\xi_i z)\,\phi(z) dz = \sum_{i=a}^{b} \delta_i \int_{z^*}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2 - \xi_i z\right) dz \\ &= \sum_{i=a}^{b} \delta_i \exp\left(\frac{1}{2}\xi_i^2\right)\big(1 - \Phi(z^* + \xi_i)\big) = \sum_{i=a}^{b} c_i P_{s,i} \Phi(-z^* - \xi_i) \end{aligned} \tag{210}$$

using the identity

$$\int_a^b \exp\left(-\frac{\alpha}{2}x^2 - \beta x\right) dx = \frac{\sqrt{2\pi}}{\sqrt{\alpha}} \exp\left(\frac{\beta^2}{2\alpha}\right)\left(\Phi\left(b\sqrt{\alpha} + \frac{\beta}{\sqrt{\alpha}}\right) - \Phi\left(a\sqrt{\alpha} + \frac{\beta}{\sqrt{\alpha}}\right)\right) \ \ \forall\ \alpha > 0 \tag{211}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function. In the case of a receiver swaption, it differs from payer swaption only by flipping the signs of c_i's (and thus the signs of δ_i's). The same argument still applies, which gives the receiver swaption price as

$$V_{s,t,a,b}^{\mathrm{RS}} = -\sum_{i=a}^{b} \delta_i \int_{-\infty}^{z^*} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2 - \xi_i z\right) dz = -\sum_{i=a}^{b} c_i P_{s,i} \Phi(z^* + \xi_i) \tag{212}$$

This is consistent with the put-call parity in swaptions, where the underlying swap value should be the payer swaption premium minus the receiver swaption premium.

Note that the formulas (210) and (212) are applicable only if the solution z^* is unique. The argument that δ_i's are all positive (negative) up to a certain $i = k$ then all negative (positive) is a sufficient but unnecessary condition for the uniqueness of z^*. It ensures δ_i and $(\xi_k - \xi_i)$ having the same sign and therefore the monotonicity in $g(z)$. However, even if the condition was not satisfied (i.e., the δ_i's change several times of sign, so do the c_i's) and the monotonicity in $g(z)$ could not be guaranteed, there would still be a good chance to have a unique z^*, especially when the sizes of irregular δ_i's are reasonably small [33]. Nevertheless, if the z^* is however not unique, the exercise domain of an option will be a union of disjoint intervals rather than a single interval, calculation of the integral must then be done by numerical integration methods.

6.3.2. Market Value of Swaption

The market value of a swaption is often quoted in lognormal volatility or normal volatility. The latter is often in favor due to the prevailing low or even negative levels of sovereign interest rates after 2008 financial crisis. Knowing that a swaption is actually a contingent claim on swap rate, we may price the payer swaption in the market by

$$V_{s,t,a,b}^{\mathrm{PS}} = A_s^{a,b}\mathbb{E}_s^{a,b}\left[\left(S_t^{a,b} - K\right)^+\right], \qquad S_t^{a,b} = \frac{P_{t,a} - P_{t,b}}{A_t^{a,b}}, \qquad A_t^{a,b} = \sum_{i=a+1}^{b} \tau_i P_{t,i} \tag{213}$$

The swap rate $S_t^{a,b}$ for $s \le t \le t_a$ is a martingale under the swap measure $\mathbb{Q}^{a,b}$ with annuity $A_t^{a,b}$ as the numeraire. If we assume $S_t^{a,b}$ is a lognormal martingale, we can compute its value by *Black* formula using the market quoted lognormal volatility ς_{LN} with $v = \varsigma_{LN}^2(t-s)$ for $t = t_a$

$$\begin{aligned} &V_{s,t,a,b}^{\mathrm{PS}} = A_s^{a,b}\mathfrak{B}\left(K, S_s^{a,b}, v, 1\right) \\ &\mathfrak{B}(K,F,v,\omega) = \omega F\Phi(\omega d^+) - \omega K\Phi(\omega d^-), \qquad d^+ = \frac{1}{\sqrt{v}}\log\frac{F}{K} + \frac{\sqrt{v}}{2}, \qquad d^- = d^+ - \sqrt{v} \end{aligned} \tag{214}$$

On the other hand, if assuming $S_t^{a,b}$ is a normal martingale, we can compute its value using the market quoted normal volatility ς_N with $v = \varsigma_N^2(t-s)$

$$V_{s,t,a,b}^{\mathrm{PS}} = A_s^{a,b}\mathbb{E}_s^{a,b}\left[\left(S_t^{a,b} - K\right)^+\right] = A_s^{a,b}\mathbb{E}_s^{a,b}\left[\left(S_s^{a,b} + Z\sqrt{v} - K\right)^+\right] = A_s^{a,b}\mathfrak{N}\left(K, S_s^{a,b}, v, 1\right) \tag{215}$$

by the *Bachelier* formula $\mathfrak{N}(K,F,v,\omega)$

$$\mathfrak{N}(K,F,v,\omega) = \omega(F-K)\Phi\left(\frac{\omega(F-K)}{\sqrt{v}}\right) + \sqrt{v}\phi\left(\frac{\omega(F-K)}{\sqrt{v}}\right) \tag{216}$$

6.4. Transition Probability Density Function

To calibrate the general Dupire local volatility, we must know the joint distribution of the FX spot and both rates in order to estimate the $\gamma_{t,K}$ in (187). It is worth mentioning that zero correlation parameters in the model (193) do not necessarily lead to vanishing expectation in the numerators of (187). This is because that these correlation parameters only characterize the dependency structure among instantaneous changes of the stochastic drivers of the state variables. Obviously, the FX spot has explicit dependency on both of the short rates through its drift term.

Let us define a function $h(t,X,\hat{x},x)$, which can be regarded as the discounted (i.e., numeraire adjusted) transition probability density function characterized by the SDE (193)

$$h(t,X,\hat{x},x)=D_{s,t}p(t,X,\hat{x},x)=D_{s,t}p(t,X_t,\hat{x}_t,x_t|s,X_s,\hat{x}_s,x_s),\qquad D_{s,t}=\exp\left(-\int_s^t r_u du\right) \tag{217}$$

where $D_{s,t}$ is the discount factor in domestic currency. The $p(t,X,\hat{x},x)=p(t,X_t,\hat{x}_t,x_t|s,X_s,\hat{x}_s,x_s)$ is the transition density under risk neutral measure, which has full knowledge of terminal distribution of the state variables, and its evolution is governed by the *Fokker-Planck* equation (10). Noting that $D_{s,t}$ is not a function of any of the time t state variables, we can write

$$\frac{\partial h}{\partial t}=-r_t h+D_{s,t}\frac{\partial p}{\partial t} \tag{218}$$

Further expanding the $\partial p/\partial t$ term by (10) gives the forward PDE of h

$$\begin{aligned}\frac{\partial h}{\partial t}=-r_t h-&\frac{\partial\left(\left(\lambda_t-\hat{\lambda}_t-\frac{1}{2}\mathcal{g}_{t,X}^2\right)h\right)}{\partial X}+\frac{\partial\left(\left(\rho_t^{01}\hat{\varsigma}_t\mathcal{g}_{t,X}+\hat{\mu}_t\hat{x}\right)h\right)}{\partial\hat{x}}+\frac{\partial(\mu_t x h)}{\partial x}+\frac{1}{2}\frac{\partial^2\left(\mathcal{g}_{t,X}^2 h\right)}{\partial X^2}\\&+\frac{1}{2}\frac{\partial^2(\hat{\varsigma}_t^2 h)}{\partial\hat{x}^2}+\frac{1}{2}\frac{\partial^2(\varsigma_t^2 h)}{\partial x^2}+\frac{\partial^2\left(\rho_t^{01}\hat{\varsigma}_t\mathcal{g}_{t,X}h\right)}{\partial X\partial\hat{x}}+\frac{\partial^2\left(\rho_t^{02}\varsigma_t\mathcal{g}_{t,X}h\right)}{\partial X\partial x}\\&+\frac{\partial^2(\rho_t^{12}\varsigma_t\hat{\varsigma}_t h)}{\partial\hat{x}\partial x}\end{aligned} \tag{219}$$

with initial condition $\lim_{t\to s}h(t,X,\hat{x},x)$ being a 3D Dirac delta function. For numerical solution, we may approximate the initial condition at $t=s+\tau$ for a small time interval τ by a 3D Gaussian density function with zero correlations

$$\begin{aligned}&\lim_{t\to s}h(t,X,\hat{x},x)\approx P_{s,t}\phi(X|\mathbb{E}[X_t],\mathbb{V}[X_t])\phi(\hat{x}|\mathbb{E}[\hat{x}_t],\mathbb{V}[\hat{x}_t])\phi(x|\mathbb{E}[x_t],\mathbb{V}[x_t])\\&\phi(x|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),\qquad\mathbb{E}[X_t]=-\frac{1}{2}\ell^2\tau,\qquad\mathbb{V}[X_t]=\ell^2\tau,\\&\mathbb{E}[\hat{x}_t]=-\rho^{01}\hat{\varsigma}\ell\tau,\qquad\mathbb{V}[\hat{x}_t]=\hat{\varsigma}^2\tau,\qquad\mathbb{E}[x_t]=0,\qquad\mathbb{V}[x_t]=\varsigma^2\tau\end{aligned} \tag{220}$$

where ℓ is the classic Dupire local volatility. The coupling of the discount factor and the density function in h allows us to compute present value of an asset under domestic risk neutral measure simply by a 3D integration once we know the h

$$\mathbb{E}_s\left[D_{s,t}V(t,X,\hat{x},x)\right] = \int_{\mathbb{R}}\int_{\mathbb{R}}\int_{\mathbb{R}} V(t,X,\hat{x},x)h(t,X,\hat{x},x)dX\,d\hat{x}\,dx \tag{221}$$

6.5. Model Calibration by Forward PDE

The calibration procedure is summarized as follows:

1. For the first time step, the initial density $h(t_0)$ is taken to be a 3D Dirac delta function (approximated by the 3D Gaussian density) and we assume $\mathscr{g}(t_0,X) = \ell(t_0,X)$ (i.e., using classic Dupire local volatility for the general Dupire local volatility)
2. For each time step from t_i to t_{i+1}, we evolve the density one time step forward from $h(t_i)$ to $h(t_{i+1})$ by (219), using previously calculated $\mathscr{g}(t_i,X)$
3. Use the resulted $h(t_{i+1})$ to compute the volatility adjustment $\gamma(t_{i+1},X)$ in (187), specifically we evaluate the numerator and the denominator using the $h(t_{i+1})$.
4. Use the computed volatility adjustment $\gamma(t_{i+1},X)$ along with the classic Dupire local volatility $\ell(t_{i+1},X)$ to compute the general Dupire local volatility $\mathscr{g}(t_{i+1},X)$ for the time interval from t_{i+1} to t_{i+2}
5. Repeat steps 2 to 4 until the density function has been evolved all the way to the maturity

However, there are two difficulties in the above procedure. Firstly, we are going to solve numerically with an extremely peaked Dirac delta density for the very first time step from t_0 to t_1. A solution to this problem is that we may skip solving the PDE and approximate the $h(t_1)$ directly to be the 3D Gaussian density using $\ell(t_0,X)$ along with other parameters for the state variables. Secondly, when we estimate $\gamma(t_{i+1},X)$, both numerator and denominator in (187) vanish at wings of volatility smile. We must design a cutoff point beyond which the volatility adjustment can be safely ignored. To minimize the numerical instability, we take the following steps

1. We can estimate the numerator in (187) using the call or the put expectation at different strike levels, e.g.,

$$c = \mathbb{E}_s\left[D_{s,t}\left(\hat{\lambda}_t \mathcal{X}_t - \lambda_t K\right)\Theta(\mathcal{X}_t - K)\right] = \int_k^\infty \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\hat{\lambda}_t \mathcal{X}_t - \lambda_t K\right) h(t, X, \hat{x}, x) d\hat{x}\, dx\, dX$$

$$p = \mathbb{E}_s\left[D_{s,t}\left(\lambda_t K - \hat{\lambda}_t \mathcal{X}_t\right)\Theta(K - \mathcal{X}_t)\right] = \int_{-\infty}^k \int_{\mathbb{R}} \int_{\mathbb{R}} \left(\lambda_t K - \hat{\lambda}_t \mathcal{X}_t\right) h(t, X, \hat{x}, x) d\hat{x}\, dx\, dX$$

When the strike level is below the ATM strike, we take the expectation from the put, otherwise take the expectation from the call.

2. The denominator in (187) is associated with the second derivatives of option prices with respect to strike. As (188) shows, they are merely the discounted transition density at the strike level. Basically, there could be 3 ways to estimate the denominator:

   a. Use the formula provided in (159), (160) or (161), which is the same denominator term when estimating classic Dupire local volatility along with a Black-Scholes vega

   b. Use finite difference to approximate the second derivatives

   c. Use numerical integration of the density function h, as in (188)

$$\frac{\partial^2 \mathcal{C}_{t,K}}{\partial K^2} = \frac{\partial^2 \mathcal{P}_{t,K}}{\partial K^2} = \mathbb{E}_s\left[D_{s,t}\delta(\mathcal{X}_t - K)\right] = P_{s,t} p_{t,K} = P_{s,t} p_{t,X} \frac{dX}{dK} = \frac{h(t,X)}{K}$$

$$= \frac{1}{K} \int_{\mathbb{R}} \int_{\mathbb{R}} h(t, X, \hat{x}, x) d\hat{x}\, dx$$

We are in favor of the 3rd method. Not only is it easy to estimate (given that we already know $h(t, X, \hat{x}, x)$ function), but also it helps to offset the numerical noise in the numerator, which is also estimated from the $h(t, X, \hat{x}, x)$ function.

6.5.1. Numerical Solution of Forward PDE

Lastly, we elaborate a bit about the method for solving the PDE. For the sake of relatively easy implementation, we seek to solve the forward PDE (219) numerically using *Marchuk-Yanenko* locally one dimensional (LOD) method[1] [34] [35]. In this method, the PDE (219) breaks into three sub-PDEs, one for each direction. However, the cross terms prevent this from occurring and they are handled explicitly in the X step

$$\begin{aligned}
\frac{\partial h}{\partial t} &= -r_t h - \frac{\partial\left(\left(\lambda_t - \hat{\lambda}_t - \frac{\mathcal{g}_{t,X}^2}{2}\right)h\right)}{\partial X} + \frac{\partial^2\left(\frac{\mathcal{g}_{t,X}^2}{2}h\right)}{\partial X^2} + \frac{\partial^2\left(\rho_t^{01}\hat{\varsigma}_t \mathcal{g}_{t,X} h\right)}{\partial X \partial \hat{x}} + \frac{\partial^2\left(\rho_t^{02}\varsigma_t \mathcal{g}_{t,X} h\right)}{\partial X \partial x} \\
&\quad + \frac{\partial^2\left(\rho_t^{12}\varsigma_t\hat{\varsigma}_t h\right)}{\partial \hat{x}\partial x} \\
\frac{\partial h}{\partial t} &= \frac{\partial\left(\left(\rho_t^{01}\hat{\varsigma}_t \mathcal{g}_{t,X} + \hat{\mu}_t \hat{x}\right)h\right)}{\partial \hat{x}} + \frac{\partial^2\left(\frac{\hat{\varsigma}_t^2}{2}h\right)}{\partial \hat{x}^2} \\
\frac{\partial h}{\partial t} &= \frac{\partial(\mu_t x h)}{\partial x} + \frac{\partial^2\left(\frac{\varsigma_t^2}{2}h\right)}{\partial x^2}
\end{aligned} \tag{222}$$

For the X step, we may transform the PDE into

$$\frac{\partial h}{\partial t} = (-r_t - T_X + C)h, \qquad T_X = \left(\lambda_t - \hat{\lambda}_t\right)\frac{\partial}{\partial X} - \left(\frac{\partial}{\partial X} + \frac{\partial^2}{\partial X^2}\right)\left(\frac{\mathcal{g}_{t,X}^2}{2}\,\cdot\right) \tag{223}$$

with the cross-term operator

$$\begin{aligned}
C &= \frac{\partial^2\left(\rho_t^{01}\hat{\varsigma}_t \mathcal{g}_{t,X}\,\cdot\right)}{\partial \hat{x}\partial X} + \frac{\partial^2\left(\rho_t^{02}\varsigma_t \mathcal{g}_{t,X}\,\cdot\right)}{\partial x \partial X} + \frac{\partial^2\left(\rho_t^{12}\varsigma_t\hat{\varsigma}_t\,\cdot\right)}{\partial x \partial \hat{x}} \\
&= \rho_t^{01}\hat{\varsigma}_t \frac{\partial}{\partial \hat{x}}\frac{\partial\left(\mathcal{g}_{t,X}\,\cdot\right)}{\partial X} + \rho_t^{02}\varsigma_t \frac{\partial}{\partial x}\frac{\partial\left(\mathcal{g}_{t,X}\,\cdot\right)}{\partial X} + \rho_t^{12}\varsigma_t \frac{\partial}{\partial x}\hat{\varsigma}_t\frac{\partial}{\partial \hat{x}} \\
&= \rho_t^{01}\mathcal{D}_X\mathcal{D}_{\hat{x}} + \rho_t^{02}\mathcal{D}_X\mathcal{D}_x + \rho_t^{12}\mathcal{D}_{\hat{x}}\mathcal{D}_x = \mathcal{D}_X\left(\rho_t^{01}\mathcal{D}_{\hat{x}} + \rho_t^{02}\mathcal{D}_x\right) + \rho_t^{12}\mathcal{D}_{\hat{x}}\mathcal{D}_x \\
\mathcal{D}_X &= \frac{\partial}{\partial X}\left(\mathcal{g}_{t,X}\,\cdot\right), \qquad \mathcal{D}_{\hat{x}} = \hat{\varsigma}_t\frac{\partial}{\partial \hat{x}}, \qquad \mathcal{D}_x = \varsigma_t\frac{\partial}{\partial x}
\end{aligned} \tag{224}$$

Let $\tau = t_{i+1} - t_i$ be the time step size, we may evolve the PDE in time implicitly by

$$\frac{h_{i+1} - h_i}{\tau} = -r h_{i+1} - T_X h_{i+1} + c \Longrightarrow (I + r\tau + T_X\tau)h_{i+1} = h_i + \tau c \tag{225}$$

[1] A brief introduction to finite difference method can be found in my notes "*Introduction to Interest Rate Models*", which can be obtained from https://modelmania.github.io/main/

For the $\hat{x}$ step, we do it similarly

$$
\begin{aligned}
&\frac{\partial h}{\partial t} = T_{\hat{x}} h, \qquad T_{\hat{x}} = \rho_t^{01} \hat{\varsigma}_t \vartheta_{t,X} \frac{\partial}{\partial \hat{x}} + \hat{\mu}_t \frac{\partial(\hat{x}\,\cdot)}{\partial \hat{x}} + \frac{\hat{\varsigma}_t^2}{2} \frac{\partial^2}{\partial \hat{x}^2} \\
&h_{i+1} - h_i = T_{\hat{x}} \tau h_{i+1} \Longrightarrow (I - T_{\hat{x}} \tau) h_{i+1} = h_i
\end{aligned}
\tag{226}
$$

and for the x step, we have

$$
\begin{aligned}
&\frac{\partial h}{\partial t} = T_x h, \qquad T_x = \mu_t \frac{\partial(x\,\cdot)}{\partial x} + \frac{\varsigma_t^2}{2} \frac{\partial^2}{\partial x^2} \\
&h_{i+1} - h_i = T_x \tau h_{i+1} \Longrightarrow (I - T_x \tau) h_{i+1} = h_i
\end{aligned}
\tag{227}
$$

For each time step, t_i to t_{i+1}, we must loop through the X step, $\hat{x}$ step and x step sequentially. The result from previous (spatial) step will be used as starting point for the next to evolve over the same time interval. For example, we first take the result of the previous time step and use it as the starting point for the X step. The result of the X step is then used as the starting point for the $\hat{x}$ step as we evolve again over the same time interval. Finally, the output from the $\hat{x}$ step is then used as the starting point of the x step. Once we have evolved all three variables from t_i to t_{i+1}, we compute the integrals for the volatility adjustment.

6.6. Pricing by Backward PDE

The pricing is done through solving a backward PDE subject to proper boundary conditions

$$
\begin{aligned}
\frac{\partial V}{\partial t} = r_t V - \left(\lambda_t - \hat{\lambda}_t - \frac{\vartheta_{t,X}^2}{2}\right) \frac{\partial V}{\partial X} + \left(\rho_t^{01} \hat{\varsigma}_t \vartheta_{t,X} + \hat{\mu}_t \hat{x}\right) \frac{\partial V}{\partial \hat{x}} + \mu_t x \frac{\partial V}{\partial x} - \frac{\vartheta_{t,X}^2}{2} \frac{\partial^2 V}{\partial X^2} - \frac{\hat{\varsigma}_t^2}{2} \frac{\partial^2 V}{\partial \hat{x}^2} \\
- \frac{\varsigma_t^2}{2} \frac{\partial^2 V}{\partial x^2} - \rho_t^{01} \hat{\varsigma}_t \vartheta_{t,X} \frac{\partial^2 V}{\partial X \partial \hat{x}} - \rho_t^{02} \varsigma_t \vartheta_{t,X} \frac{\partial^2 V}{\partial X \partial x} - \rho_t^{12} \varsigma_t \hat{\varsigma}_t \frac{\partial^2 V}{\partial x \partial \hat{x}}
\end{aligned}
\tag{228}
$$

where $V = V(t, X, \hat{x}, x)$ with a terminal condition being the payoff of an asset upon maturity. Again, we use *Marchuk-Yanenko* locally one dimensional (LOD) method for the solution. The PDE breaks down into the following three sub-PDEs

$$\begin{aligned}
\frac{\partial V}{\partial t} &= r_t V - \left(\lambda_t - \hat{\lambda}_t - \frac{\mathcal{g}_{t,X}^2}{2}\right)\frac{\partial V}{\partial X} - \frac{\mathcal{g}_{t,X}^2}{2}\frac{\partial^2 V}{\partial X^2} - \rho_t^{01}\hat{\varsigma}_t \mathcal{g}_{t,X}\frac{\partial^2 V}{\partial X \partial \hat{x}} - \rho_t^{02}\varsigma_t \mathcal{g}_{t,X}\frac{\partial^2 V}{\partial X \partial x} \\
&\qquad - \rho_t^{12}\varsigma_t\hat{\varsigma}_t \frac{\partial^2 V}{\partial x \partial \hat{x}} \\
\frac{\partial V}{\partial t} &= \left(\rho_t^{01}\hat{\varsigma}_t \mathcal{g}_{t,X} + \hat{\mu}_t \hat{x}\right)\frac{\partial V}{\partial \hat{x}} - \frac{\hat{\varsigma}_t^2}{2}\frac{\partial^2 V}{\partial \hat{x}^2} \\
\frac{\partial V}{\partial t} &= \mu_t x \frac{\partial V}{\partial x} - \frac{\varsigma_t^2}{2}\frac{\partial^2 V}{\partial x^2}
\end{aligned} \tag{229}$$

For the X step, we evolve the PDE backwards using an implicit scheme, as described below

$$\begin{aligned}
&\frac{\partial V}{\partial t} = T_X V - c, \qquad T_X = r_t + \frac{\mathcal{g}_{t,X}^2}{2}\left(\frac{\partial}{\partial X} - \frac{\partial^2}{\partial X^2}\right) - \left(\lambda_t - \hat{\lambda}_t\right)\frac{\partial}{\partial X} \\
&\frac{V_i - V_{i-1}}{\tau} = T_X V_{i-1} - c \Longrightarrow (I + T_X \tau) V_{i-1} = V_i + \tau c
\end{aligned} \tag{230}$$

where the cross term is estimated explicitly in the X step

$$\begin{aligned}
c &= \rho_t^{01}\hat{\varsigma}_t \mathcal{g}_{t,X}\frac{\partial^2 V}{\partial X \partial \hat{x}} + \rho_t^{02}\varsigma_t \mathcal{g}_{t,X}\frac{\partial^2 V}{\partial X \partial x} + \rho_t^{12}\varsigma_t\hat{\varsigma}_t \frac{\partial^2 V}{\partial x \partial \hat{x}} \\
&= \left(\rho_t^{01}\hat{\varsigma}_t \frac{\partial}{\partial \hat{x}} \mathcal{g}_{t,X}\frac{\partial}{\partial X} + \rho_t^{02}\varsigma_t \frac{\partial}{\partial x}\mathcal{g}_{t,X}\frac{\partial}{\partial X} + \rho_t^{12}\varsigma_t \frac{\partial}{\partial x}\hat{\varsigma}_t \frac{\partial}{\partial \hat{x}}\right) V \\
&= \left(\frac{\rho_t^{12}}{\rho_t^{01}\rho_t^{02}}\mathcal{D}_{\hat{x}}\mathcal{D}_x + \mathcal{D}_X(\mathcal{D}_x + \mathcal{D}_{\hat{x}})\right) V
\end{aligned} \tag{231}$$

$$\mathcal{D}_X = \mathcal{g}_{t,X}\frac{\partial}{\partial X}, \qquad \mathcal{D}_{\hat{x}} = \rho_t^{01}\hat{\varsigma}_t \frac{\partial}{\partial \hat{x}}, \qquad \mathcal{D}_x = \rho_t^{02}\varsigma_t \frac{\partial}{\partial x}$$

For the $\hat{x}$ step, we have

$$\begin{aligned}
&\frac{\partial V}{\partial t} = T_{\hat{x}} V, \qquad T_{\hat{x}} = \left(\rho_t^{01}\hat{\varsigma}_t \mathcal{g}_{t,X} + \hat{\mu}_t \hat{x}\right)\frac{\partial}{\partial \hat{x}} - \frac{\hat{\varsigma}_t^2}{2}\frac{\partial^2}{\partial \hat{x}^2} \\
&V_i - V_{i-1} = T_{\hat{x}}\tau V_{i-1} \Longrightarrow (I + T_{\hat{x}}\tau) V_{i-1} = V_i
\end{aligned} \tag{232}$$

and for the x step, we have

$$\frac{\partial V}{\partial t} = T_x V, \qquad T_x = \mu_t x \frac{\partial}{\partial x} - \frac{\varsigma_t^2}{2}\frac{\partial^2}{\partial x^2} \tag{233}$$

$$V_i - V_{i-1} = T_x \tau V_{i-1} \Longrightarrow (I + T_x \tau) V_{i-1} = V_i$$

Similar to what we have done for the forward PDE, for each time step, t_i to t_{i-1}, we loop through the X step, $\hat{x}$ step and x step each individually. The result from previous (spatial) step will be used as starting point for the next to evolve over the same time interval.

For barrier options, it is often more convenient to use grid in log spot moneyness, e.g., $\tilde{X}_t = \log(\mathcal{X}_t/\mathcal{X}_s)$, for simple treatment of boundaries. The corresponding spot dynamics in (193) is then transformed into

$$d\tilde{X}_t = \left(r_t - \hat{r}_t - \frac{1}{2}\tilde{\mathscr{g}}_{t,\tilde{X}}^2\right)dt + \tilde{\mathscr{g}}_{t,\tilde{X}} dW_{0,t}, \tag{234}$$

The local volatility component $\tilde{\mathscr{g}}(t,\tilde{X})$ can be interpolated from the calibrated $\mathscr{g}(t,X)$ surface through conversion in (158), that is

$$\tilde{\mathscr{g}}(t,\tilde{X}) = \mathscr{g}(t,X) = \mathscr{g}\left(t, \log\frac{\mathcal{X}_t}{F_{s,t}}\right) = \mathscr{g}\left(t, \log\frac{\mathcal{X}_t}{\mathcal{X}_s} + \log\frac{\mathcal{X}_s}{F_{s,t}}\right) = \mathscr{g}\left(t, \tilde{X} + \log\frac{\mathcal{X}_s}{F_{s,t}}\right) \tag{235}$$

Let $\tilde{V} = \tilde{V}(t,\tilde{X},\hat{x},x)$ be the corresponding value of a discounted contingent claim. The change of variable from X to $\tilde{X}$ would produce a PDE that differs from (228) only by the term $\lambda_t - \hat{\lambda}_t$ that replaces $r_t - \hat{r}_t$ in coefficient of $\partial\tilde{V}/\partial\tilde{X}$. The PDE can be solved in the same manner as stated previously.

7. STOCHASTIC LOCAL VOLATILITY: ORNSTEIN-UHLENBECK DUPIRE

Ornstein-Uhlenbeck Dupire model, a full-fledged 2D stochastic local volatility model. The stochastic volatility component is modeled as an Exponential Ornstein-Uhlenbeck process with mean reversion. Calibration of the local volatility component is based on Gyöngy theorem by 2D forward induction. (Available upon request …)

8. STOCHASTIC LOCAL VOLATILITY: MARKOV CHAIN ORNSTEIN-UHLENBECK DUPIRE

Markov Chain Ornstein-Uhlenbeck Dupire model. A simplified version of the OUDupire model. The stochastic volatility component is assumed to be driven by an independent discrete Markov chain process, equivalent to the Exponential Ornstein-Uhlenbeck process in OUDupire model. This simplification greatly improves computational efficiency as it avoids solving 2D PDE in calibration. Instead, we solve 1D PDE at different state levels and remix the states in a forward induction manner. This model is superior to the MixedDupire model as it possesses mean reversion dynamics, which is more realistic to describe the forward volatility dynamics. (Available upon request …)

9. STOCHASTIC LOCAL VOLATILITY: MIXED DUPIRE

Mixed Dupire local volatility (MixedDupire) model and its variants have been widely used in the industry to price first-generation FX exotics. It is a simple stochastic local volatility model. Its stochastic volatility component is modeled as an initial random shock that calibrates to volatility smiles. The local volatility component is calibrated based on Gyöngy theorem by solving 1D PDE at different state levels of the initial shock and remixing the states in a forward induction manner. (Available upon request …)

10. STOCHASTIC LOCAL VOLATILITY WITH STOCHASTIC RATES: MIXED GENERAL DUPIRE

Multi-state general Dupire local volatility model. This is a stochastic local volatility model with stochastic rates, extended from the GeneralDupire model and the MixedDupire model. The stochastic volatility component is modeled the same as in the MixedDupire model, which is driven by an initial random shock. Both stochastic rates are modeled as Hull-White 1 factor short rate model, the same as in GeneralDupire model. The local volatility component is calibrated based on Gyöngy theorem, by solving

3D PDE at different state levels of the initial shock and remixing the states in a forward induction manner. (Available upon request …)

11. Local Volatility: PDE by Finite Difference Method

In this chapter, we will present a PDE based local volatility model, in which the local volatility surface is constructed as a 2-D function that is piecewise constant in maturity and piecewise linear in log-moneyness (for equity) or delta (for FX). Due to great similarity between FX and equity processes, our interest lies primarily in the context of equity derivatives, the conclusions and formulas drawn from our discussion here are in general applicable to FX products with minor changes. In contrast to the traditional way to construct the local volatility by estimating highly sensitive and numerically unstable partial derivatives in Dupire formulas, this method relies heavily on solving forward PDE's to calibrate a parametrized local volatility surface to vanilla option prices in a bootstrapping manner. Once the local volatility surface is calibrated, the backward PDE can then be used to price exotic options (e.g., barrier options) that are in consistent with the market observed implied volatility surface.

Before proceeding to the PDE's, it is important to have an overview of the date conventions for equity and equity options. The date conventions for FX products are defined in a similar manner.

11.1. Date Conventions of Equity and Equity Option

The diagram illustrates the date definitions for an equity and its associated option. The quantities appeared in the diagram are listed in Table 5.

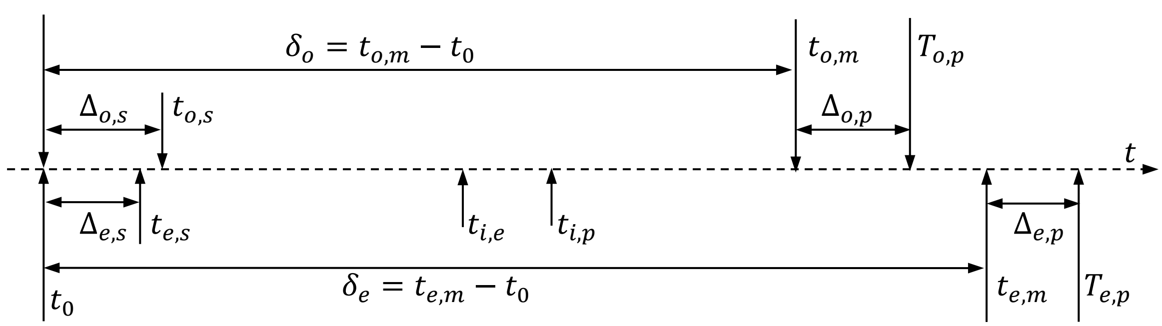


Table 5. Dates of Equities and Options

attribute	symbol	description	remark/example
trade date	t_0	on which the equity/option is traded	today
equity spot lag	$\Delta_{e,s}$	equity premium settlement lag	3D
equity spot date	$t_{e,s}$	on which the equity premium is settled	$t_{e,s} = t_0 \oplus \Delta_{e,s}$
equity maturity date	$t_{e,m}$	equity maturity date	$t_{e,m} = t_0 \oplus 1Y$
equity pay lag[1]	$\Delta_{e,p}$	lag between $t_{e,m}$ and $t_{e,p}$	e.g., same as $\Delta_{e,s}$
equity pay date	$t_{e,p}$	on which the equity payoff is settled	$t_{e,p} = t_{e,m} \oplus \Delta_{e,p}$
i-th dividend	θ_i	dividend payment amount	
i-th ex- div. date	$t_{i,e}$	ex-dividend date	
i-th div. pay date	$t_{i,p}$	dividend pay date	
option spot lag	$\Delta_{o,s}$	option premium settlement lag	2D
option spot date	$t_{o,s}$	on which the option is settled	$t_{o,s} = t_0 \oplus \Delta_{o,s}$
option maturity date	$t_{o,m}$	option maturity date	$t_{o,m} = t_0 \oplus 1Y$
option pay lag	$\Delta_{o,p}$	lag between $t_{o,m}$ and $t_{o,p}$	e.g., same as $\Delta_{o,s}$
option pay date	$t_{o,p}$	on which the equity payoff is settled	$t_{o,p} = t_{o,m} \oplus \Delta_{o,p}$
day rolling	$\oplus$	rolling with convention "following"	Following
calendar		defining business days and holidays	US / UK / HK

As most of the quantities are self-explanatory, our discussion focuses more on the treatment of equity dividends.

11.2. Deterministic Dividends

In our example, we can assume both the short rate and the dividend rate are deterministic and continuous, e.g., time-dependent r_t and q_t as in (122). the equity forward in this case can be calculated by

$$F(t_0, t_{e,m}) = X(t_0)\frac{P_q(t_{e,s}, t_{e,p})}{P_r(t_{e,s}, t_{e,p})} \qquad \text{where}$$
$$P_q(s,t) = \exp\left(-\int_s^t q_u du\right), \qquad P_r(s,t) = \exp\left(-\int_s^t r_u du\right) \tag{236}$$

In a more realistic implementation, we may assume the underlying equity issues a series of discrete dividends with fixed amounts in a foreseeable future. It is obvious that the equity spot still follows the SDE (122) with $q_t = 0$ in between two adjacent ex-dividend dates (There is discontinuity in spot process

[1] Equity settlement delay

on ex-dividend dates that demands special treatment. This will be discussed in detail in due course). With fixed dividends, the equity forward becomes

$$F\big(t_0, t_{e,m}\big) = \frac{X(t_0) - \sum_i \theta_i P_r\big(t_{e,s}, t_{i,p}\big)}{P_r\big(t_{e,s}, t_{e,p}\big)} \quad \text{for} \quad t_0 < t_{i,e} \le t_{e,m} \tag{237}$$

where θ_i is the fixed amount of the i-th dividend issued on ex-dividend date $t_{i,e}$.

Discrete dividend can also be modeled as proportional dividend. It assumes that at each ex-dividend date, the dividend payment will result in a price drop in equity spot proportional to the spot level. For example, the equity spot before and after the dividend fall has the relationship

$$X\big(t_{i,e} + \Delta\big) = X\big(t_{i,e} - \Delta\big)(1 - \eta_i) \tag{238}$$

where Δ denotes an infinitesimal amount of time and η_i the proportional dividend rate at ex-dividend date $t_{i,e}$. By this relationship, we can write the equity forward as

$$F\big(t_0, t_{e,m}\big) = X(t_0) \frac{\prod_i (1 - \eta_i)}{P_r\big(t_{e,s}, t_{e,p}\big)} \quad \text{for} \quad t_0 < t_{i,e} \le t_{e,m} \tag{239}$$

Sometimes it is often more convenient to approximate the fixed dividends by proportional dividends. The conversion can be achieved by equating the equity forward in (237) and (239), such that

$$\prod_i (1 - \eta_i) = 1 - \frac{1}{X(t_0)} \sum_i \theta_i P_r\big(t_{e,s}, t_{i,p}\big) \quad \text{for} \quad t_0 < t_{i,e} \le t_{e,m} \tag{240}$$

The proportional dividend η_i can then be bootstrapped from a series of fixed dividends θ_i starting from the first ex-dividend date.

11.3. Forward PDE

In the following, our derivation relies on the spot process X_s given in (122) and its variants. Specifically we may write the SDE (122) in terms of log-spot $z_t = \log X_t$ or centered log-spot $z_t = \log\big(X_t / F_{s,t}\big)$

$$dz_t = \left(\mu_t - \frac{1}{2}\varsigma(t,z)^2\right) dt + \varsigma(t,z) dW_t \quad \text{and} \quad dz_t = -\frac{1}{2}\varsigma(t,z)^2 dt + \varsigma(t,z) dW_t \tag{241}$$

where $\varsigma(t,z)$ and $\varsigma(t,z)$ are the local volatility function in z and z, respectively.

Let's denote the forward time variable by t for $s < t$ and use centered log-spot z_t for the process. Given that $z_s = 0$, the value of a normalized undiscounted call can be defined as

$$V_{t,k|s,z} = \frac{C_{t,k|s,z}}{F_{s,t}} = \frac{\mathbb{E}[(X_t - K)^+|s, X_s]}{F_{s,t}} \tag{242}$$

where $k = \log(K/F_{s,t})$ is the strike in log-moneyness (as in (135)). Let $\varsigma_{t,k}$ be the local volatility function in k, which is equivalent to $\varsigma_{t,K}$, we can derive forward PDE for $V_{t,k|s,z}$ from (144)

$$\begin{aligned} \frac{\varsigma_{t,k}^2}{2} &= \frac{F_{s,t}\dfrac{\partial V_{t,k|s,z}}{\partial t} + \mu_t F_{s,t} V_{t,k|s,z} - \mu_t C_{t,k|s,z}}{F_{s,t}\dfrac{\partial^2 V_{t,k|s,z}}{\partial k^2} - F_{s,t}\dfrac{\partial V_{t,k|s,z}}{\partial k}} = \frac{\dfrac{\partial V_{t,k|s,z}}{\partial t}}{\dfrac{\partial^2 V_{t,k|s,z}}{\partial k^2} - \dfrac{\partial V_{t,k|s,z}}{\partial k}} \\ \Rightarrow \frac{\partial V_{t,k|s,z}}{\partial t} &= \frac{\varsigma_{t,k}^2}{2}\left(\frac{\partial^2 V_{t,k|s,z}}{\partial k^2} - \frac{\partial V_{t,k|s,z}}{\partial k}\right) \end{aligned} \tag{243}$$

with initial condition

$$V_{s,k|s,z} = \frac{C_{s,k|s,z}}{F_{s,s}} = \frac{\mathbb{E}\left[\left(X_s - F_{s,s}e^k\right)^+ \middle| s, X_s\right]}{F_{s,s}} = (1 - e^k)^+ \tag{244}$$

using the partial derivatives

$$\frac{\partial V_{t,k|s,z}}{\partial t} = \frac{1}{F_{s,t}}\frac{\partial C_{t,k|s,z}}{\partial t} - \mu_t V_{t,k|s,z}, \quad \frac{\partial V_{t,k|s,z}}{\partial k} = \frac{1}{F_{s,t}}\frac{\partial C_{t,k|s,z}}{\partial k}, \quad \frac{\partial^2 V_{t,k|s,z}}{\partial k^2} = \frac{1}{F_{s,t}}\frac{\partial^2 C_{t,k|s,z}}{\partial k^2} \tag{245}$$

The PDE (243) appears drift-less and provides more robust calibration stability at low volatility and/or high drift due to the "transparency" of drift in the PDE.

11.3.1. <u>Treatment of Deterministic Dividends</u>

A (discrete) dividend pay-out will typically result in a drop in equity price on the ex-dividend date. Suppose that time t is the ex-dividend date, the no-arbitrage condition states that at τ the time right after the ex-dividend date (e.g., the difference between t and τ can be infinitesimal), we must have

$$X_\tau = X_t - \theta_t \tag{246}$$

where θ_t is the *value* of dividend issued at t (note that in a rigorous setup the value must take into account the discounting effect due to dividend payment delay). Since a forward is expectation of spot under risk neutral measure[1], we may write

$$F_{s,\tau} = \mathbb{E}_s[X_\tau] = \mathbb{E}_s[X_t - \theta_t] = F_{s,t} - \mathbb{E}_s[\theta_t] \tag{247}$$

Under the assumption that θ_t is a fixed amount, it reads

$$F_{s,\tau} = F_{s,t} - \theta_t \tag{248}$$

In our finite difference method, the spatial grid for log-moneyness k is assumed uniform such that $k_i - k_{i-1}$ is constant for all i. Dividend payment causes discontinuity in the underlying spot. Evolving the forward PDE (243) from initial time s produces a state vector $V_{t,k|s,z}$ at time t. Immediately after the issuance of dividend at time τ, the spot and forward drop the same θ_t amount and hence the state vector $V_{\tau,k|s,z}$ must be realigned to reflect the dividend fall. This can be done using the option no-arbitrage condition, such that

$$\begin{gathered} C_{\tau,k|s,z} = \mathbb{E}_s[(X_\tau - K)^+] = \mathbb{E}_s\left[\left(X_t - \theta_t - F_{s,\tau}e^k\right)^+\right] = \mathbb{E}_s\left[\left(X_t - F_{s,t}e^{\hat{k}}\right)^+\right] = C_{t,\hat{k}|s,z} \\ \text{where} \quad \hat{k} = \log\frac{F_{s,\tau}e^k + \theta_t}{F_{s,t}} \end{gathered} \tag{249}$$

Subsequently we can use $\hat{k}$ to interpolate from the $V_{t,k|s,z}$ state vector and transform the interpolated value to form $V_{\tau,k|s,z}$ vector by

$$V_{\tau,k|s,z} = \frac{C_{\tau,k|s,z}}{F_{s,\tau}} = \frac{C_{t,\hat{k}|s,z}}{F_{s,t}}\frac{F_{s,t}}{F_{s,\tau}} = \frac{F_{s,t}}{F_{s,\tau}}V_{t,\hat{k}|s,z} \tag{250}$$

If the dividend is proportional, we must have spot price $X_\tau = X_t(1-\eta_t)$ for a rate η_t and hence forward price $F_{s,\tau} = F_{s,t}(1-\eta_t)$ before and after the dividend fall. Because we can show that

[1] Strictly speaking, a forward on time T spot is an expectation of the spot under T-forward measure, i.e. $F_{t,T} = \mathbb{E}_t^T[X_T]$. However since the interest rate is assumed deterministic, the T-forward measure coincides with the risk neutral measure.

$$V_{\tau,k|s,z} = \frac{\mathbb{E}_s[(X_\tau - K)^+]}{F_{s,\tau}} = \frac{(1-\eta_t)\mathbb{E}_s\left[\left(X_t - F_{s,t}e^k\right)^+\right]}{F_{s,t}(1-\eta_t)} = \frac{\mathbb{E}_s\left[\left(X_t - F_{s,t}e^k\right)^+\right]}{F_{s,t}} = V_{t,k|s,z} \tag{251}$$

the state vector remains unchanged before and after the issuance of dividend.

With continuous dividend q_t, the realignment of state vector is unnecessary because there is no discontinuity in equity spot.

11.4. Backward PDE

Again we assume the spot follows the SDE (122). Without loss of generality, let's denote $G(X_T|K)$ an arbitrary payoff function with parameter K, whose value is contingent on X_T at maturity T. One example of such function would be the payoff function of a call option: $G(X_T|K) = (X_T - K)^+$. Let $U_{t,x|T,K}$ be the expectation of the function $G(X_T|K)$ at time t with spatial variable $x = X_t$, which can be written as

$$U_{t,x|T,K} = \mathbb{E}[G(X_T|K)|t,x] = \int_{\mathbb{R}} G(y|K) p_{T,y|t,x} dy \tag{252}$$

where the transition probability $p_{T,y|t,x}$ follows the *Kolmogorov backward equation* (20)

$$\frac{\partial p_{T,y|t,x}}{\partial t} = \mu_t x \frac{\partial p_{T,y|t,x}}{\partial x} + \frac{\varsigma_{t,x}^2 x^2}{2} \frac{\partial^2 p_{T,y|t,x}}{\partial x^2} \tag{253}$$

In turn, we can derive the backward PDE for the $U_{t,x|T,K}$ such that

$$\begin{aligned}\frac{\partial U_{t,x|T,K}}{\partial t} &= \int_{\mathbb{R}} G(y|K) \frac{\partial p_{T,y|t,x}}{\partial t} dy = -\int_{\mathbb{R}} G(y|K) \left(\mu_t x \frac{\partial p_{T,y|t,x}}{\partial x} + \frac{\varsigma_{t,x}^2 x^2}{2} \frac{\partial^2 p_{T,y|t,x}}{\partial x^2}\right) dy \\ &= -\mu_t x \frac{\partial U_{t,x|T,K}}{\partial x} - \frac{\varsigma_{t,x}^2 x^2}{2} \frac{\partial^2 U_{t,x|T,K}}{\partial x^2}\end{aligned} \tag{254}$$

with terminal condition

$$U_{T,x|T,K} = G(x|K) \tag{255}$$

11.4.1. PDE in Centered Log-spot

Assuming the spatial variable is $z_t = \log(x/F_{s,t})$ at time t, we may write $U_{t,z|T,k}$ in the (t,z)-plane equivalent to $U_{t,x|T,K}$. The backward PDE (254) can then be transformed into

$$\frac{\partial U_{t,z|T,k}}{\partial t} = -\frac{\varsigma_{t,z}^2}{2}\left(\frac{\partial^2 U_{t,z|T,k}}{\partial z^2} - \frac{\partial U_{t,z|T,k}}{\partial z}\right) \tag{256}$$

with terminal condition

$$U_{T,z|T,k} = G\left(F_{s,T}e^z \middle| F_{s,T}e^k\right) \tag{257}$$

by using the following partial derivatives derived from the chain rule

$$\begin{aligned}
&\frac{\partial z}{\partial x} = \frac{1}{x}, \qquad \frac{\partial z}{\partial t} = -\mu_t, \qquad \frac{\partial U_{t,x|T,K}}{\partial t} = \frac{\partial U_{t,z|T,k}}{\partial t} + \frac{\partial U_{t,z|T,k}}{\partial z}\frac{\partial z}{\partial t} = \frac{\partial U_{t,z|T,k}}{\partial t} - \mu_t \frac{\partial U_{t,z|T,k}}{\partial z} \\
&\frac{\partial U_{t,x|T,K}}{\partial x} = \frac{\partial U_{t,z|T,k}}{\partial z}\frac{\partial z}{\partial x} = \frac{1}{x}\frac{\partial U_{t,z|T,k}}{\partial z}, \qquad \frac{\partial^2 U_{t,x|T,K}}{\partial x^2} = \frac{1}{x^2}\left(\frac{\partial^2 U_{t,z|T,k}}{\partial z^2} - \frac{\partial U_{t,z|T,k}}{\partial z}\right)
\end{aligned} \tag{258}$$

11.4.1.1. *Treatment of Deterministic Dividends*

With fixed dividend θ_t, we have

$$X_\tau = X_t - \theta_t \qquad \text{and} \qquad F_{s,\tau} = F_{s,t} - \theta_t \tag{259}$$

The no arbitrage condition shows that for the spatial grid z

$$\begin{aligned}
&U_{t,z|T,k} = \mathbb{E}\left[G\left(X_T|F_{s,T}e^k\right)\middle|t, F_{s,t}e^z\right] = \mathbb{E}\left[G\left(X_T|F_{s,T}e^k\right)\middle|\tau, F_{s,t}e^z - \theta_t\right] \\
&= \mathbb{E}\left[G\left(X_T|F_{s,T}e^k\right)\middle|\tau, F_{s,\tau}e^{\hat{z}}\right] = U_{\tau,\hat{z}|T,k} \qquad \text{where} \qquad \hat{z} = \log\frac{F_{s,t}e^z - \theta_t}{F_{s,\tau}}
\end{aligned} \tag{260}$$

It is likely that if z is sufficiently small (e.g., at lower boundary of spatial grid) we may end up with $F_{s,t}e^z - \theta_t < 0$, which makes the $\hat{z}$ not well defined. A solution is to floor it to a small positive number, e.g., taking $\max\left(10^{-10},\ F_{s,t}e^z - \theta_t\right)$. This is valid because equity spot must be positive and the $U_{\tau,\hat{z}|T,k}$ flattens as $\hat{z}$ goes to negative infinity. After the special treatment, we can use the $\hat{z}$ to interpolate from the $U_{t,z|T,k}$ state vector and convert the interpolated value into vector $U_{\tau,z|T,k}$.

With proportional dividend, the conclusion drawn for forward PDE still applies here and the state vector remains unchanged before and after the dividend fall. With continuous dividend, the realignment of state vector is unnecessary because there is no discontinuity in equity spot.

11.4.1.2. *Vanilla Call*

Due to the duality between the forward and backward PDE, it is evident that vanilla calls (or puts) must admit the identity: $U_{s,z|T,k} = V_{T,k|s,z} F_{s,T}$, where $U_{s,z|T,k}$ is the undiscounted call solved from backward PDE (256) and $V_{T,k|s,z}$ the normalized undiscounted call solved from forward PDE (243). This relationship can be used to check the correctness of implementation of the numerical engines of forward and backward PDE.

11.4.2. PDE in Log-spot

For pricing some exotic options, e.g., barrier options, it is more convenient to use log-spot $z = \log x$ as the spatial variable. Similarly we can define $\mathscr{k} = \log K$. Let us denote the (discounted) price of a derivative product by

$$Q_{t,z|T,\mathscr{k}} = P_{t,T} U_{t,z|T,\mathscr{k}} = \mathbb{E}\left[D_{t,T} G(X_T | e^k) | t, e^z\right] \tag{261}$$

By taking into account the discount factor, it must follow the following backward PDE

$$\begin{aligned}
\frac{\partial Q_{t,z|T,\mathscr{k}}}{\partial t} &= r_t Q_{t,z|T,\mathscr{k}} + P_{t,T} \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial t} \\
&= r_t Q_{t,z|T,\mathscr{k}} + P_{t,T} \left(-\mu_t x \frac{1}{x} \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} - \frac{\varsigma_{t,z}^2}{2} \frac{1}{x^2} \left(\frac{\partial^2 U_{t,z|T,\mathscr{k}}}{\partial z^2} - \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} \right) \right) \\
&= -\frac{\varsigma_{t,z}^2}{2} \frac{\partial^2 Q_{t,z|T,\mathscr{k}}}{\partial z^2} + \left(\frac{\varsigma_{t,z}^2}{2} - \mu_t \right) \frac{\partial Q_{t,z|T,\mathscr{k}}}{\partial z} + r_t Q_{t,z|T,\mathscr{k}}
\end{aligned} \tag{262}$$

where the partial derivatives below have been used

$$\begin{aligned}
&\frac{\partial z}{\partial x} = \frac{1}{x}, \qquad \frac{\partial z}{\partial t} = 0, \qquad \frac{\partial U_{t,x|T,K}}{\partial t} = \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial t} + \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} \frac{\partial z}{\partial t} = \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial t} \\
&\frac{\partial U_{t,x|T,K}}{\partial x} = \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial t} \frac{\partial t}{\partial x} + \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} \frac{\partial z}{\partial x} = \frac{1}{x} \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} \\
&\frac{\partial^2 U_{t,x|T,K}}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{1}{x} \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} \right) = -\frac{1}{x^2} \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} + \frac{1}{x} \frac{\partial^2 U_{t,z|T,\mathscr{k}}}{\partial z \partial t} \frac{\partial t}{\partial x} + \frac{1}{x} \frac{\partial^2 U_{t,z|T,\mathscr{k}}}{\partial z^2} \frac{\partial z}{\partial x} \\
&\qquad\qquad\quad = \frac{1}{x^2} \left(\frac{\partial^2 U_{t,z|T,\mathscr{k}}}{\partial z^2} - \frac{\partial U_{t,z|T,\mathscr{k}}}{\partial z} \right)
\end{aligned} \tag{263}$$

11.4.2.1. *Treatment of Deterministic Dividends*

With fixed dividend θ_t, the no arbitrage condition states that

$$\begin{aligned} Q_{t,z|T,k} &= \mathbb{E}\left[D_{t,T}G\left(X_T|e^{k}\right)\middle|t, e^{z}\right] = \mathbb{E}\left[D_{\tau,T}G\left(X_T|e^{k}\right)\middle|\tau, e^{z}-\theta_t\right] \\ &= \mathbb{E}\left[D_{\tau,T}G\left(X_T|e^{k}\right)\middle|\tau, e^{\hat{z}}\right] = Q_{\tau,\hat{z}|T,k} \qquad \text{where} \qquad \hat{z} = \log(e^{z}-\theta_t) \end{aligned} \tag{264}$$

Again, extremely small z may result in $\hat{z}$ that is not well defined, we may floor the difference $e^{z}-\theta_t$ to a small positive number, e.g., taking $\max(10^{-10},\ e^{z}-\theta_t)$. The vector $Q_{t,z|T,k}$ can then be interpolated from the known $Q_{\tau,z|T,k}$ using the $\hat{z}$.

With proportional dividend η_t, again the no arbitrage condition shows

$$\begin{aligned} Q_{t,z|T,k} &= \mathbb{E}\left[D_{t,T}G\left(X_T|e^{k}\right)\middle|t, e^{z}\right] = \mathbb{E}\left[D_{\tau,T}G\left(X_T|e^{k}\right)\middle|\tau, e^{z}(1-\eta_t)\right] \\ &= \mathbb{E}\left[D_{\tau,T}G\left(X_T|e^{k}\right)\middle|\tau, e^{\hat{z}}\right] = Q_{\tau,\hat{z}|T,k} \qquad \text{where} \qquad \hat{z} = z + \log(1-\eta_t) \end{aligned} \tag{265}$$

The vector $Q_{t,z|T,k}$ can be interpolated from the $Q_{\tau,z|T,k}$ using the $\hat{z}$.

With continuous dividend, the realignment of state vector is unnecessary because there is no discontinuity in equity spot.

11.5. Local Volatility Surface

This section is devoted to discussing the construction of local volatility surface $\varsigma(t,k)$. There are various ways to define the local volatility surface. The one that we would like to discuss is a 2-D function that is piecewise constant in maturity t and piecewise linear in log-moneyness $k = \log\left(K/F_{s,t}\right)$ (or in delta for FX). The volatility surface comprises a series of volatility smiles $\sigma_j(k)$ for maturity $s < t_1 < \cdots < t_j < \cdots < t_m = T$. At each maturity t_j, volatility smile $\sigma_j(k)$ is constructed by linear interpolation between log-moneyness pillars $k_i = \log\left(K_i/F_{s,t}\right)$ for strikes $K_1 < \cdots < K_i < \cdots < K_n$ and flat extrapolation where the volatility values at k_1 and k_n are used for all $k < k_1$ and $k > k_n$, respectively. The smile $\sigma_j(k)$ constructed at t_j is assumed to remain constant over time for any t between the two adjacent maturities $t_{j-1} < t \le t_j$.

Calibration of the local volatility surface is conducted in a bootstrapping manner starting from the shortest maturity t_1. It is done by solving the *forward* PDE such that the local volatility surface is able to reproduce the vanilla call prices at the prescribed log-moneyness pillars k_i for each of the maturities t_j. The PDE can be solved using finite difference method[1] on a uniform grid defined on log-moneyness k that extends to ± 5 standard deviations of the underlying spot. The choice of boundary condition has little impact to the solutions of vanilla option prices because at ± 5 standard deviations the transition probability becomes negligibly small. Our application uses linearity boundary condition for its simplicity. To allow a higher tolerance to market data input and smoother calibration process, the objective function may include a penalty term to suppress unfavorable concavity of a local volatility smile. Again, there can be many ways to define the objective function as well as the penalty function. In this essay, we will only focus on the simplest objective (e.g., at maturity t_j): the least square minimization of vanilla call prices

$$\underset{\sigma_j(k_i)}{\operatorname{argmin}} \sum_{i=1}^{n} \left(U^{\mathrm{BS}}_{s,z|t_j,k_i} - U^{\mathrm{PDE}}_{s,z|t_j,k_i} \right)^2 \tag{266}$$

where $U_{s,z|T,k}$ is the normalized undiscounted call price defined in (242), the superscript "BS" denotes the theoretical price by Black-Scholes model and the "PDE" denotes the numerical value by forward PDE. Note that without a penalty term, the minimization can lead to an exact solution given a proper[2] implied volatility surface.

11.6. Barrier Option Pricing

In contrast to the calibration, the pricing of a barrier option relies on the *backward* PDE (262) in line with proper terminal condition (i.e., payoff function) and boundary conditions defined by the characteristics of the barrier option. Barrier options often demand a spatial grid defined on log-spot $z = \log X_t$, which allows an easier fit of time-invariant barrier (e.g., with European or American type of

[1] A brief introduction to finite difference method can be found in my notes "*Introduction to Interest Rate Models*", which can be downloaded from https://modelmania.github.io/main/

[2] A proper implied volatility surface should well behave and admit no arbitrage

observation window) into the domain. For example, an up-and-out barrier option would be priced on a domain with upper bound at the barrier level b where Dirichlet boundary condition is applied (the lower bound and its boundary condition remain the same as for vanilla options).

REFERENCES

1. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley-Finance, 2011, pp.82
2. Bjork, T., *The Pedestrian's Guide to Local Time*, 2015
   https://arxiv.org/pdf/1512.08912.pdf
3. Wikipedia Page: https://en.wikipedia.org/wiki/Local_time_(mathematics)
4. Gyöngy, I., *Mimicking the one-dimensional marginal distributions of processes having an Ito differential*, Probability Theory and Related Fields, 71:501-516, 1986
5. Austing, P., *Smile Pricing Explained*, Palgrave McMillan, 2014, chapter 9.1
6. Piterbarg, V., *Markovian Projection Method for Volatility Calibration*, SSRN, 2006
   https://papers.ssrn.com/sol3/papers.cfm?abstract_id=906473
7. Haugh, M., *Local Volatility, Stochastic Volatility and Jump-Diffusion Models*, Course Slides, 2013
   https://martin-haugh.github.io/files/ContinuousFE/LocalStochasticJumps.pdf
8. Kotzé, A.; Oosthuizen, R.; Pindza, E., (2015) : *Implied and local volatility surfaces for South African index and foreign exchange options*, Journal of Risk and Financial Management, ISSN 1911-8074, MDPI, Basel, Vol. 8, Iss. 1, pp. 43-82
   https://www.econstor.eu/bitstream/10419/178554/1/jrfm-08-00043.pdf
9. Gatheral, J., *The Volatility Surface: A Practitioner's Guide*, Wiley-Finance, 2006, pp. 8
10. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley Finance, 2011. pp. 41-43
11. Reiswich, D. and Wystup, U., *A Guide to FX Options Quoting Conventions*, The Journal of Derivatives, Winter 2010, Vol. 18, No. 2: pp. 58-68
    https://www.researchgate.net/publication/275905055_A_Guide_to_FX_Options_Quoting_Conventions
12. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley Finance, 2011. pp. 43
13. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley Finance, 2011. Chapter 3
14. Wystup, U., *FX Options and Structured Products, 2nd Ed*, Wiley Finance, 2017. pp. 17
15. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley Finance, 2011. pp. 48-49
16. Reiswich, D. and Wystup, U., *FX Volatility Smile Construction*, Working Paper, 2010. pp. 9-11
    Online: https://www.econstor.eu/bitstream/10419/40186/1/613825101.pdf
17. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley Finance, 2011. pp. 52
18. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley Finance, 2011. pp. 59
19. Gatheral, J., *A parsimonious arbitrage-free implied volatility parameterization*, 2004

Online resource: http://faculty.baruch.cuny.edu/jgatheral/madrid2004.pdf

20. Gatheral, J., *Arbitrage-free SVI volatility surfaces*, Quantitative Finance, Vol 14 Issue 1, 2014, pp. 59-71
https://arxiv.org/pdf/1204.0646.pdf

21. Lee, R., *The Moment Formula for Implied Volatility at Extreme Strikes*, Mathematical Finance, Vol. 14, No. 3, pp. 469-480, July 2004. (Available at SSRN: http://ssrn.com/abstract=558885)

22. Zeliade 2012 - Quasi-Explicit Calibration of Gatheral's SVI model

23. Aurell 2014 MS Thesis - The SVI implied volatility model and its calibration

24. Ferhati 2020 - Robust Calibration For SVI Model Arbitrage Free

25. Clark, I., *Foreign Exchange Option Pricing - A Practitioner's Guide*, Wiley Finance, 2011. pp. 69-70

26. Austing, P., *Smile Pricing Explained*, Palgrave McMillan, 2014, chapter 7.1, pp. 72-73

27. Gatheral, J., *The Volatility Surface: A Practitioner's Guide*, Wiley-Finance, 2006, chapter 8, pp. 102-103

28. Deelstra, G. and Rayée, G., *Local Volatility Pricing Models for Long-dated FX Derivatives*, 2012
https://arxiv.org/pdf/1204.0633

29. Online resource: http://itf.fys.kuleuven.be/~nikos/papers/lect4_localvol.pdf

30. Gatheral, J., *The Volatility Surface: A Practitioner's Guide*, Wiley-Finance, 2006, pp. 13-14

31. Andersen and Piterbarg 2010 - *Interest Rate Modeling*, ch. 10.1.2.3 page 418-419

32. Henrard, M., *Explicit Bond Option and Swaption Formula in Heath-Jarrow-Morton One Factor Model*, International Journal of Theoretical and Applied Finance, Vol. 6, No. 1, pp. 57-72, 2003
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=434860

33. Online Resource:
http://forums.opengamma.com/t/question-on-swaption-price-in-quant-research-paper-hull-white-one-factor-model-results-and-implementation/366

34. Marchuk, G. I., *Splitting and alternating direction methods*, Handbook of Numerical Analysis Vol. I, Ciarlet, P. and Lions, J. Editors, North-Holland, Amsterdam, 1990, 197–462.

35. Glowinski, R., *Finite Element Methods for Incompressible Viscous Flow*, Handbook of Numerical Analysis, Vol. IX, Ciarlet, P. and Lions, J. Editors, North-Holland, Amsterdam, 2003, Chapter 6, pp. 433-563