OPTION PRICING WITH GAUSSIAN FUNCTIONALS

R.S

Abstract. SABR approximation does not guarantee arbitrage free prices, as a consequence recent researches have focused on developing SABR models that yield positive density function. Instead of improving SABR models, the paper presents a model that prices European options while being consistent with the static and dynamic market smile.

Introduction

The construction of financial models compatible with the market smile is an abiding challenge. Traditional models such as Black76 and Bachelier never manage to replicate the smile as neither the implied Normal volatility smile, nor the implied Black smile are flat. Moreover, the procedure of delta hedging under these models is a hazardous operation due to the implied volatilities changing following a rate movement.

Local volatility models were introduced to tackle the smile replication issue, with Dupire proposing a method to build the market local volatility from a continuum of option prices across times and strikes. In practice, the poorness of observable prices requires building artificial prices to calibrate the local volatility model, hence the need of price interpolator. More importantly, the non-parametric form of the local volatility implies that the smile dynamics (i.e the smile movement when the rate changes) is entirely incorporated in the market prices. The market has witnessed different market regimes where the Black/Bachelier at-the money implied volatility of several European options were roughly invariant with respect to rate movement, a proper model must offer the ability of controlling this dynamics, also known as ATM backbone.

Stochastic volatility models such as the infamous SABR models possess extra risk factors that can generate several rate distributions that fit the market smile, as well as a handle on the ATM backbone. The classic SABR models is defined with 4 parameters : α controls the level of the smile, ρ the slope, ν the convexity, and β the ATM backbone. We can note that the stochastic volatility parameters that may affect the smile dynamics - mainly the factor $\rho\nu$ when β is fixed - are not exogenous parameters and decided by one snapshot of the market smile. For instance, if one observes that a 50% correlation between rate and volatility yields a coherent smile dynamics, whereas today's smile requires a -20% correlation, then the latter must be used and the parameter β deals with the dynamics.

Hagan developed a closed-form formula that connects the SABR parameters to the implied normal volatilities, it is accurate up to a certain expiry and strike, beyond these values the approximated formula generates prices far from the actual model, and worse, the density function can be negative especially around the absorbing point $F = 0$ for $0 < \beta \leq 1$. Moreover, the SABR model cannot calibrate to extreme strikes options while matching liquid points. A satisfactory option model must allow the control of the smile wings, which can be crucial for CMS pricing by replication.

The recent literature has offered several methodologies to address the presence of arbitrage in the SABR model. Doust and Hagan reconstruct the density function by guaranteeing its positive sign, ZABR/Balland constructs the local volatility version of SABR using respectively Gyongy theorem and short-maturity expansion, with the latter model calibrated using a one-time step Dupire forward PDE. The construction of a local volatility version of a stochastic volatility model is not a trivial task due to the calculation of its harmonic average, in addition, the issue of accuracy with the actual model for long-term maturity is still present.

In this paper, instead of introducing an n^{th} way of fixing SABR issues, we develop a new model that addresses all the issues encounters by its predecessors. It generates arbitrage-free prices, all type of volatility smiles : control of the level, the slope, the convexity, the left wing for zero-strike floors, the right wing for high strike calls and for the extreme right for the CMS pricing. The model must handle negative rates and offers a full control of the ATM backbone. Finally, the computational time should be minimum.

European option pricing

The financial market is defined by the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, N)$. The undiscounted call option on the reference rate R_T with strike K and expiry T is defined as follows

$$c(t,K) = E^N[(R_T - K)^+ |\mathcal{F}_t] \tag{1}$$

The forward rate F_t tied to the market rate R_T is

$$F_t = E^N[R_T|\mathcal{F}_t], \quad 0 \leq t \leq T \tag{2}$$

where F_0 is generated by the linear rate market. In practice, R_T can be a δ-tenor Libor rate fixed at time T and N is the $(T+\delta)$-forward measure, or a swap rate fixed at time T and N its natural annuity measure. By construction, the process $(F_t)_{0\leq t\leq T}$ is a $\mathcal{F}$-martingale, and is often modeled using a driftless stochastic process with initial value F_0. Traditionally, Black (3) and Bachelier(Normal)(4) models are used, they respectively assume a log-normal and normal dynamics. Nowadays, the models are widely used as quotation format, as a consequence of their dynamics being inconsistent with the market observations.

$$dF_t = \sigma F_t dW_t \tag{3}$$

$$dF_t = \sigma dW_t \tag{4}$$

Practitioners have replaced (3) and (4) with more sophisticated models such as the infamous SABR model.

Terminal distribution. Using the definition of the forward rate F_t (2) , (1) can be written as follows

$$c(t,K) = E^N[(F_T - K)^+ |\mathcal{F}_t] \tag{5}$$

Similarly to linear instrument pricing, the whole trajectory of F is not needed to price options, only the distribution of F_T at time T must be known. In this paper, F_T is modeled directly without introducing a stochastic differential equation.
The distribution of R_T is written as a function of n independent random variables $\{U_i\}_{i=1}^n$

$$R_T = G(U_1, U_2, ..., U_n) \tag{6}$$

Under this model, the price of a call option at time 0 is

$$c(0,K) = \int_{\mathbb{R}^n} [G(u_1, u_2, ..., u_n) - K]^+ \prod_{i=1}^n f_{U_i}(u_i) du_i \tag{7}$$

with f_{U_i} the density function of the variable U_i.
The function G is chosen such as it replicates the market smile, the market backbone , last but not least, it must verify

$$F_0 = \int_{\mathbb{R}^n} G(u_1, u_2, ..., u_n) \prod_{i=1}^n f_{U_i}(u_i) du_i \tag{8}$$

The SDE (3) and (4) are particular case of (6) with $n = 1$, respectively, through the functions G_{LN} and G_N defined as

$$\begin{aligned} G_{LN} &: u \mapsto F_0 e^{-\frac{1}{2}\sigma^2 T + \sigma\sqrt{T}u} \\ G_N &: u \mapsto F_0 + \sigma\sqrt{T}u \\ f_{U_1} &: u \mapsto \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \end{aligned} \tag{9}$$

The Gaussian functionals in (9) are not sufficiently malleable to match the market distribution. In the next section, we introduce a function G directed by a system of random ordinary differential equations.

Random ordinary differential equations

Let $(\Omega, \mathcal{F}, N)$ denotes a probability space, and let $h : \mathbb{R}^d \times \mathbb{R}^n \longrightarrow \mathbb{R}^d$ be a continuous function.
A random ordinary differential equation in $\mathbb{R}^d$ is defined as follows,

$$
\begin{aligned}
\frac{dX}{dt} &= h(X, U(\omega)), \quad X \in \mathbb{R}^d \\
U(\omega) &:= (U_1(\omega), ..., U_n(\omega)) \\
\omega &\in \Omega
\end{aligned}
\tag{10}
$$

where $\{U_i\}_{i=1}^n$ are independent random variables.
For almost every realization $\omega \in \Omega$, the RODE (11) is also an autonomous ordinary differential equation (ODE)

$$
\frac{dX}{dt} = H_\omega(X) := h(X, U(\omega)), \quad X \in \mathbb{R}^d
\tag{11}
$$

For the sack of clarity, we can remove the realizations of nil probability from Ω, hence (12) is valid for all realizations. The existence and uniqueness theorems for ODEs apply to (12) for a given realization ω.
For a given realization ω, we define $R_T(\omega) = g(X_T(\omega))$, where $X_T(\omega)$ is the solution of (12) at time T, the function $g : \mathbb{R}^d \longrightarrow \mathbb{R}$ verifies

$$
E(g(X_T)) < \infty.
$$

Multi-dimensional Gaussian RODE. Henceforth, each element of $\{U_i\}_{i=1}^n$ is a standard normal variable, and we will investigate the following family of Gaussian RODE

$$
\begin{aligned}
\frac{dX(t, U(\omega))}{dt} &= \sigma_X[X(t, U(\omega)), U(\omega)]\gamma(U(\omega)) \\
X(0, u(\omega)) &= x_0 \quad a.s \\
\omega &\in \Omega \\
0 \leq t &\leq T
\end{aligned}
\tag{12}
$$

with $\sigma_X : \mathbb{R}^d \times \mathbb{R}^n \longrightarrow \mathbb{R}^d \times \mathbb{R}^d$ strictly positive and differentiable, the map $\gamma : \mathbb{R}^n \longrightarrow \mathbb{R}^d$ is linear.

Two-factor models. In the latter model, $n = d = 2$ and the following subfamily of (12) is defined

$$
\begin{aligned}
\frac{dX(t, U)}{dt} &= \alpha(t, U)C(X(t, U))\gamma_1(U) \\
\frac{d\alpha(t, U)}{dt} &= D(\alpha(t, U))\gamma_2(U) \\
\gamma_1(U) &:= \frac{\rho U_1 + \sqrt{1 - \rho^2}U_2}{\sqrt{T}} \\
\gamma_2(U) &:= \frac{U_1}{\sqrt{T}} \\
X(0, U) &= x_0 \quad a.s \\
\alpha(0, U) &= \alpha > 0 \quad a.s \\
0 \leq t &\leq T
\end{aligned}
\tag{13}
$$

with the functions C and D strictly positive and differentiable.

For each $\omega \in \Omega$, the RODE (13) also verifies

$$\frac{\partial X(T,u)}{\partial u_1} = \alpha(T,u_1)C(X(T,u))\rho\sqrt{T} \tag{14}$$

$$\frac{\partial X(T,u)}{\partial u_2} = \alpha(T,u)C(X(T,u))\sqrt{1-\rho^2}\sqrt{T} \tag{15}$$

$$\frac{d\alpha(T,u_1)}{du_1} = \nu D(\alpha(T,u_1))\sqrt{T} \tag{16}$$

The new formulation (14)-(16) can be proved by using the differential form H

$$\begin{aligned} dH(x,y) &= \frac{1}{yC(x)}dx + \frac{1}{D(y)}dy \\ (x,y) &\in \mathbb{R}\times\mathbb{R}_+^* \end{aligned} \tag{17}$$

Solving the ODEs (14)-(16) is more efficient than solving (13). Furthermore, it is worth noting that $(u_1,u_2) \mapsto X(T,u)$ is a strictly increasing function.

European call option price. Before using (7) to price European options, the realizations of U must be generated. The set of realizations are by nature concentrated in 0, however, the ODEs (14)-(16) are solved instead of (13) , hence small increments of u are requested.
For the sake of clarity, only one state grid $\{u_i\}_{i=1}^N$ is used to define realizations (u_1,u_2), with

$$u_i = u_{min} + \frac{u_{max}-u_{min}}{N} i$$

N is chosen large enough to comply with the ODE scheme requirement, it uses the step $s := \frac{u_{max}-u_{min}}{N}$.
Similarly, $u_{min} := -u_{max}$ and u_{max} are large enough, verifying

$$\mathbb{P}(U_1 < u_{min}) = \mathbb{P}(U_1 > u_{max}) = \epsilon$$

with ϵ very small.
Finally, the realization $u_1 = u_2 = 0$ must be included in the state grid as $X(T,0) = x_0 := F_0$, which is the initial condition of the ODE (13).
The ODEs (14)-(16) are solved with multiple steps:

(1) The current state is $u_1 = 0$, and the ODE (14) is solved in two steps for $u_2 \in]0,u_{max}]$ and $u_2 \in [u_{min},0[$.
(2) Use (15) and (16), to get $X(T,u_1,0)$ and $\alpha(T,u_1)$ for either states $u_1 = \pm s$.
(3) Update the state u_1, repeat (1).

As stated in the previous section, the function $u \mapsto X(T,u)$ is strictly increasing, therefore a smooth and monotonic interpolator can be used to create the missing points between the ones generated by the ODE solver.
The function G introduced in (6) is then defined as

$$\begin{aligned} G : (u_1,u_2) &\mapsto \hat{G}(u_1,u_2) + F_0 - \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \phi(u)e^{-\frac{1}{2}u^2}du \\ \hat{G} : (u_1,u_2) &\mapsto \Psi_{2D}\left(\{u_i\}_{i=1}^N, \{u_j\}_{j=1}^N, \{X(T,u_i,u_j)\}_{i,j=1}^N, (u_1,u_2)\right) \\ \phi : u &\mapsto \Psi_{1D}\left(\{u_i\}_{i=1}^N, \{\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\zeta_{u_i}(v)e^{-\frac{1}{2}v^2}dv\}_{i=1}^N, u\right) \\ \zeta_u : v &\mapsto \Psi_{1D}\left(\{u_i\}_{i=1}^N, \{X(T,u_i,u)\}_{i=1}^N, v\right) \end{aligned}$$

where the function $\Psi_{1D} : \mathbb{R}^{2N}\times\mathbb{R} \longrightarrow \mathbb{R}$ and $\Psi_{2D} : \mathbb{R}^{3N}\times\mathbb{R}^2 \longrightarrow \mathbb{R}$ are monotonic interpolator.
The complex and crucial part of constructing this model is defining the function σ_X(or C and α) in a way that it matches both static and dynamics market smile.

Construction of the volatility function σ_X

The objective of this section is to introduce a volatility function σ_X that shares some of the desirable SABR features, that is controlling the skew, convexity and the ATM backbone. Hereinafter, the function σ_X will be either a sole function of the forward rate level, similarly to local volatility models, or a function of the rate and an extra normal variable.

SABR models. The Hagan et al SABR models for a forward rate F_t are introduced as follows :

$$\begin{aligned} dF_t &= \alpha_t C(F_t) dW_t \\ d\alpha_t &= \nu\alpha_t \left(\rho dW_t + \sqrt{1-\rho^2} dZ_t\right) \\ \alpha_0 &= \alpha \\ dW_t dZ_t &= 0 \end{aligned} \tag{18}$$

with C an arbitrary functional form. The most common choice for C is $C(F) = F^\beta$, which defines the classic SABR.

Taking a look at the SABR equation (18), we can have some intuition of the model dynamics. It is visible that the function $f(\alpha_t, F_t) := \alpha_t C(F_t)$ is the instantaneous volatility of F_t, and a Taylor-expansion around (α_0, F_0) yields :

$$f(\alpha_t, F_t) = f(\alpha_0, F_0) + \frac{\partial f}{\partial \alpha}(\alpha_t - \alpha_0) + \frac{\partial f}{\partial F}(F_t - F_0) + \mathcal{O}((\alpha_t-\alpha_0)^2 + (F_t - F_0)^2)$$

or

$$f(\alpha_t, F_t) = \alpha_0 C(F_0) + C(F_0)(\alpha_t - \alpha_0) + \alpha_0 C'(F_0)(F_t - F_0) + \mathcal{O}((\alpha_t-\alpha_0)^2 + (F_t - F_0)^2)$$

We can write α_t in terms of F_t

$$d\alpha_t = \frac{\nu\rho}{C(F_t)} dF_t + \alpha_t \nu\sqrt{1-\rho^2} dZ_t$$

Using Euler-Maruyama method, the process α_t becomes

$$\alpha_t - \alpha_0 = \frac{\nu\rho}{C(F_0)}(F_t - F_0) + \alpha_0\nu\sqrt{1-\rho^2} Z_t$$

The volatility process α_t has two orthogonal sources of randomness : the rate and pure volatility components. Plugging XX to YY,

$$f(\alpha_t, F_t) = \alpha_0 C(F_0) + (\nu\rho + \alpha_0 C'(F_0))(F_t - F_0) + (\alpha_0 C(F_0)\nu\sqrt{1-\rho^2} Z_t) + \mathcal{O}((\alpha_t-\alpha_0)^2 + (F_t - F_0)^2)$$

We define

$$\sigma_0 := \alpha_0 C(F_0)$$

σ_0 can be seen as the ATM normal volatility by freezing time to zero in (1). The first part $\sigma_0 + (\nu\rho + \alpha_0 C'(F_0))(F_t - F_0)$ of $f(\alpha_t, F_t)$ is linear in F_t, therefore the level and the skew of the smile are mainly driven by respectively σ_0 and $(\nu\rho + \frac{d\sigma_0}{dF_0})$. The convexity is handled by $(\sigma_0\nu\sqrt{1-\rho^2} Z_t) + \mathcal{O}((\alpha_t-\alpha_0)^2 + (F_t - F_0)^2)$. In this expansion, we can confirm the empirical observations that controls the smile : $\nu\rho$ controls the skew and $\nu\sqrt{1-\rho^2}$ controls the convexity. A more intuitive way of looking at the SABR model is to look at the couple $\left(\hat{\rho} := \nu\rho, \hat{\nu} := \nu\sqrt{1-\rho^2}\right)$ instead of (ρ, ν).

$$f(\alpha_t, F_t) = \sigma_0 + (\hat{\rho} + \frac{d\sigma_0}{dF_0})(F_t - F_0) + \sigma_0\hat{\nu} Z_t + \mathcal{O}((\alpha_t-\alpha_0)^2 + (F_t - F_0)^2)$$

No doubt that $(\hat{\rho}, \hat{\nu})$ alters the smile dynamics, however, one cannot use $(\hat{\rho}, \hat{\nu})$ to control it ,indeed, the sole purpose of ρ and ν is to match the smile at a given point in time, hence a reduced legitimacy of the extra stochastic factor Z.

The element that gives us control of the dynamics is the choice of the functional form $C(F)$: it controls both backbone and skew dynamics.

The first-order analysis of the general SABR model brings in several candidates for the volatility function σ_X, for a given range of F_t, we can define

$$\sigma_X(F_t, Z) = \sigma_0(1 + \hat{\nu}\sqrt{t}Z) + \left(\hat{\rho} + \frac{d\sigma_0}{dF_0}\right)(F_t - F_0) \qquad (19)$$
$$Z \sim \mathcal{N}(0,1)$$

Another candidate that only involves the level of the forward rate

$$\sigma_X(F_t) = \sigma_0 + \left(\hat{\rho} + \frac{d\sigma_0}{dF_0}\right)(F_t - F_0) + \hat{\nu}|F_t - F_0| \qquad (20)$$

The convexity term $\hat{\nu}(F_t - F_0)$ comes from the Bachelier model, as a result of

$$F_t - F_0 = \sigma_0 Z_t$$

combined with the ad-hoc assumption of 100% correlation between W_t and Z_t, in other words the only source of randomness is the forward rate itself. The absolute value is employed to guarantee that- ceteris paribus- $\sigma_X(F_t)$ is an increasing function of $\hat{\nu}$.
The parameters $\hat{\rho}$, $\hat{\nu}$ and σ_0 locally control the shape of the smile. In the model, we must define the validity domain $[F_0 - \kappa, F_0 + \kappa]$ of XXX, where κ is a positive value. A practical choice for κ is one standard deviation in terms of ATM volatility σ_{ATM}, but this choice can be harmful for short-term options. For respectively low and high strike controls, we introduce two parameters a_{min} and a_{max} that will reduce/increase the volatility function σ_X. Finally, for very high strikes, mainly for CMS replication, we use the parameter a_{CMS}.
Using formula (19) and (20), we respectively have

$$\begin{aligned}
\sigma^+ &= \sigma(F_0 + \kappa) = \sigma_0 + \left(\hat{\rho} + \frac{d\sigma_0}{dF_0} + \hat{\nu}\right)\kappa \\
\sigma^- &= \sigma(F_0 - \kappa) = \sigma_0 - \left(\hat{\rho} + \frac{d\sigma_0}{dF_0} - \hat{\nu}\right)\kappa \\
\sigma_z^+ &= \sigma(F_0 + \kappa, z) = \sigma_0(1 + \hat{\nu}z\sqrt{T}) + \left(\hat{\rho} + \frac{d\sigma_0}{dF_0}\right)\kappa \\
\sigma_z^- &= \sigma(F_0 - \kappa, z) = \sigma_0(1 + \hat{\nu}z\sqrt{T}) - \left(\hat{\rho} + \frac{d\sigma_0}{dF_0}\right)\kappa
\end{aligned}$$

We define the cutoff parameter γ that corresponds to the number of κ that leads the forward rate to the zone controlled by either a_{min} or a_{max}. The cutoff parameter γ_{cms} holds a similar role and is tied to a_{cms} Combined with formula XX, we have

$$\begin{aligned}
\sigma^{++} &= \sigma(F_0 + \gamma\kappa) = \sigma_0 + \gamma\left(\hat{\rho} + \frac{d\sigma_0}{dF_0} + (\hat{\nu} + a_{max})\right)\kappa \\
\sigma^{--} &= \sigma(F_0 - \gamma\kappa) = \sigma_0 - \gamma\left(\hat{\rho} + \frac{d\sigma_0}{dF_0} - (\hat{\nu} + a_{min})\right)\kappa \\
\sigma^{cms} &= \sigma(F_0 + \gamma_{cms}\kappa) = \sigma^{++} + a_{cms}\left(\gamma_{cms} - \gamma\right)\kappa \\
\sigma_z^{++} &= \sigma(F_0 + \gamma\kappa, z) = \sigma_0(1 + \hat{\nu}z\sqrt{T}) + \gamma\left(\hat{\rho} + \frac{d\sigma_0}{dF_0} + (\hat{\nu} + a_{max})\right)\kappa \\
\sigma_z^{--} &= \sigma(F_0 - \gamma\kappa, z) = \sigma_0(1 + \hat{\nu}z\sqrt{T}) - \gamma\left(\hat{\rho} + \frac{d\sigma_0}{dF_0} - (\hat{\nu} + a_{min})\right)\kappa \\
\sigma_z^{cms} &= \sigma(F_0 + \gamma_{cms}\kappa, z) = \sigma^{++} + a_{cms}\left(\gamma_{cms} - \gamma\right)\kappa
\end{aligned}$$

The intermediate points of the local volatility function are created by using interpolators. The three points (F_0, σ_0), $(F_0 + \sigma_{ATM}\sqrt{T}, \sigma^+)$ and $(F_0 - \sigma_{ATM}\sqrt{T}, \sigma^-)$ are interpolated using a second-order polynomials. $(F_0 + \sigma_{ATM}\sqrt{T}, \sigma^+)$ and $(F_0 + \gamma\sigma_{ATM}\sqrt{T}, \sigma^{++})$ are connected with a second-order polynomial, the first

derivative is matched on the former point. The same thing is done for the points $(F_0 - \sigma_{ATM}\sqrt{T}, \sigma^-)$ and $(F_0 - \gamma\sigma_{ATM}\sqrt{T}, \sigma^{--})$.
The points $(F_0 + \gamma\sigma_{ATM}\sqrt{T}, \sigma^{++})$ and $(F_0 + \gamma_{cms}\sigma_{ATM}\sqrt{T}, \sigma^{cms})$ are connected using linear interpolation. The local volatility function must be composed with another to ensure that it is strictly positive.

ATM backbone. Practitioners have wished for a model capable to predict the right ATM volatility when the reference rate moves, in an exogenous way. Indeed, it is quite difficult to operate at the SDE level and ultimately translate to the new and expected ATM volatility. Instead, practitioners often use methods called "shadow delta", that is using an ATM volatility backbone model on top of the existing option model.
Various form of backbone have been used in the literature (Rebonato- in terms of Black76), mainly using power functions.

$$\sigma_{ATM}(F) = aF^b$$

or

$$\sigma_{ATM}(F') = \sigma_{ATM}(F)\left(\frac{F'}{F}\right)^b$$

In the negative rate world, one should be careful when manipulating these function. In this paper, we define a backbone model as follows

$$\sigma_{ATM}(F') = \sigma_{ATM}(F)\Pi(F', F)$$

where $\Pi : \mathbb{R}^2 \to \mathbb{R}_*^+$ is the backbone function. In equation XX, one can see that the zero-order approximation of $f(\alpha_t, F_t)$ is $\alpha_0 C(F_0)$, which can also be considered as the ATM volatility, or σ_0 in our model setup. For convenience, and this without affecting the motivation of the model, we assume that

$$\frac{\partial\sigma_0}{\partial F} = \frac{\partial\sigma_{ATM}}{\partial F}$$

Numerical Results

Appendix

References