MASTER THESIS REPORT

SABR Model Extensions For Negative Rates

Author :
Najib JAMJAM

Supervisor :
Camilla Johansson Landén
Ethan REINER

November, 2016

Acknowledgements

First of all, I want to express my gratitude and heartfelt thanks to my tutor at AXA IM, Ethan REINER for sharing his knowledge and for his monitoring throughout the period of my internship. He was always available for providing helpful advices.

I also want to thank my tutor at KTH, Camilla Johansson Landén for accepting to supervise my thesis, for her insightful guidance and for giving me very constructive feedback.

I will not end without thanking Martin, Camille, Eduardo, Jean-Benoît, Maxence and Ichem, for the amazing atmosphere and for the support during that period.

Abstract

In this report, we present extensions of the SABR model to negative rates applied to the swaption market. We start by briefly presenting the classical SABR model. Then we study the Shifted, Free Boundary and Mixture SABR Models. Numerical experiments are performed on these models to assess their performance, in particular we detail the calibration process for the Mixture SABR and apply it on market data.

Sammanfattning

I den här rapporten presenterar vi utvidgningar av SABR modellen till negativa räntor på swaption marknaden. Vi börjar med att översiktligt presentera den klassiska SABR modellen. Därefter studerar vi Shifted, Free Boundary and Mixture SABR modeller. Vidare utförs numeriska experiment på dessa modeller för att utvärdera dem. Speciellt förklaras kalibreringsprocessen i mer detalj för Mixture SABR och den tillämpas sedan på marknads data.

Contents

Chapter 1

Introduction

The simplest and most commonly used model for pricing European options is the Black-Scholes model. The log-normal model assumes that the underlying asset follows a geometric Brownian motion. Under this model, there is a bijection between the option price and the volatility. Therefore when using Black-Scholes for pricing options with different strikes, one should use the same volatility. This is not coherent with market quotes, as the observed implied volatility is not a constant function of the strike, in fact we observe a volatility skew or smile.

In order to take into account this behavior, local and stochastic volatility models were introduced. They offer an arbitrage free approach to match market quotes. In this thesis, we focus on the SABR model and its extensions to negative rates applied to the swaption market.

The SABR (Stochastic Alpha Beta Rho) model was first introduced by Hagan et al. in [1], where an approximation of the Black Scholes implied volatility as a function of the maturity and strike was derived. This formula is simple to implement and its execution is almost instantaneous, hence its popularity among practitioners. Nevertheless, this approximation has some limitations as many assumptions were made in order obtain the final formula. In order to deal with such constraints many authors proposed refinements and new methods. This topic will be addressed in the first part of this thesis.

In the SABR model with $\beta \neq 0$ rates are implicitly assumed to be strictly positive. This hypothesis seemed appropriate at the time the model was first presented. Recently rates have reached negative levels. Thus an extension to negative rates has become crucial. The second part of this thesis present such extensions, starting with the Shifted SABR and ending by a detailed study of the Mixture SABR [2].

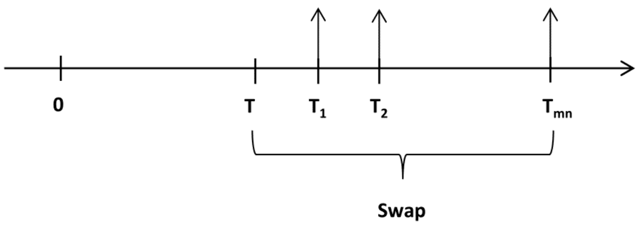
Numerical experiments are performed in the second part, as we are interested in calibrating and using the models with negative rates. The final goal is to use the Mixture SABR with market data, and assess its performance in comparison with the Shifted SABR.

Chapter 2

Swaption pricing

A swaption is an interest rate option where the holder has the right to enter an interest rate swap at a certain time in the future with a prefixed swap rate K. Let T denote the maturity of the option, and T_n the maturity of the underlying swap. The time frame $[T, T_n]$ is commonly referred to as the tenor.

We assume that each fixed payment on the swap is the fixed rate times $\tau_i L$, where L is the notional amount and τ_i is the year fraction for the payment at T_i. Therefore the cash flows are exchanged on the swap payment dates: T_1, T_2,..,T_N.



We define the swap rate F_t as the fixed rate that makes the swap having a zero value at time t.Assuming the forward swap rate at maturity is F_T, the payoff of a payer swaption is:

$$\tau_i L(F_T - K)^+ \text{ at time } T_i \in \{T_1, T_2, .., T_N\} \tag{2.1}$$

Therefore the payoff at maturity T is given by:

$$\sum_{i=1}^{N} \tau_i L P(T, T_i)\ (F_T - K)^+ \tag{2.2}$$

We define the annuity as $A(t) = \sum_{i=1}^{n} \tau_i P(t, T_i)$. It is the value of contract that pays τ_i at times $(T_i)_{i \in \{1,..,N\}}$. Using this notation the present value payoff of the swaption at time T can be written as:

$$L\ A(T)\ (F_T - K)^+ \tag{2.3}$$

Since the annuity is a traded asset with a strictly positive value, the price of the swaption can be obtained by considering the equivalent forward risk neutral measure with respect to $A(t)$.

$$V(t) = L\ A(t)\ E^A[(F_T - K)^+] \tag{2.4}$$

In this report the swap rate F_T is modeled under the annuity measure, the price of a swaption is then the price of call (or put) on F_T times the annuity. In particular E^A denotes the expectation under this measure.

Furthermore the value of the rate in a mono curve framework is given by:

$$F_t = \frac{P(t, T_0) - P(t, T_N)}{A(t)} \tag{2.5}$$

Therefore F_t is a martingale under the annuity measure and:

$$F_t = E^A[F_T] \tag{2.6}$$

This result also holds in a multi curve framework (see [3]), as the denominator doesn't change.

Example: Log-normal model

We assume that the forward rate follows a log-normal model. Based on equations 2.4 & 2.6 the price is given by Black's formula:

$$V(t) = LA(t)[F_0\mathcal{N}(d_1) - K\mathcal{N}(d_2)] \tag{2.7}$$

where

$$\begin{aligned} d_1 &= \frac{\log(\frac{F_0}{K}) - \frac{\sigma^2}{2}(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= d_1 - \sigma\sqrt{T-t} \end{aligned}$$

and $\mathcal{N}$ the standard normal distribution function. One can find further details for the derivations above in [4].

Part I

The SABR Model

Chapter 3

Hagan implied volatility approximation

The SABR model was first introduced in Hagan *et al.*[1] (2002). Since then it has been extensively used by practitioners mainly as an interpolation tool for swaptions cubes (Implied Volatility, Maturity, Tenor). Under this model, the underlying forward rate has the following dynamics:

$$dF_t = \sigma_t {F_t}^{\beta} d{W_t}^1, \quad F(0) = f \tag{3.1}$$

$$d\sigma_t = \nu \sigma_t d{W_t}^2, \quad \sigma(0) = \alpha \tag{3.2}$$

where $d{W_t}^1 d{W_t}^2 = \rho dt$. W^1, W^2 are Wiener processes and α, β, ρ are constants.
In this chapter we will first present the main result of Hagan et al.[1], then we will give some instructive elements of the proof that will enable us to understand the limits of the approximations leading to the final formula.

3.1 Main results

Hagan *et al.* presented a closed-form approximation of a Black's implied volatility as a function of the spot f, the strike K, the maturity T and the SABR parameters. Thus the price of a European option is obtained by inserting this implied volatility into a Black pricer.
The implied volatility $\sigma_B(T, K)$ is given by the following asymptotic approximation:

$$\sigma_B(T,K) = \frac{\alpha}{(fK)^{(1-\beta)/2}\left\{1 + \frac{(1-\beta)^2}{24}\log^2(f/K) + \frac{(1-\beta)^4}{1920}\log^4(f/K) + ...\right\}} \cdot \left(\frac{z}{x(z)}\right) \cdot \tag{3.3}$$
$$\left\{1 + \left[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(fK)^{1-\beta}} + \frac{1}{4}\frac{\rho\beta\nu\alpha}{(fK)^{(1-\beta)/2}} + \frac{2-3\rho^2}{24}\nu^2\right]T + ...\right.$$

Here z and $x(z)$ are defined by:

$$z = \frac{\nu}{\alpha}(fK)^{(1-\beta)/2}\log(f/K) \tag{3.4}$$

$$x(z) = \log\left\{\frac{\sqrt{1-2\rho z + z^2} + z - \rho}{1-\rho}\right\} \tag{3.5}$$

This formula (3.3) is reduced for at-the-money case $K = f$ to:

$$\sigma_{ATM} = \sigma_B(f,K) = \frac{\alpha}{f^{(1-\beta)}}\left\{1 + \left[\frac{(1-\beta)^2}{24}\frac{\alpha^2}{f^{(2-2\beta)}} + \frac{1}{4}\frac{\rho\beta\alpha\nu}{f^{(1-\beta)}} + \frac{2-3\rho^2}{24}\nu^2\right]T + ...\right. \tag{3.6}$$

These formulas are the main results in Hagan's *et al.* paper [1]. This approximation has the virtue of being straightforward to implement, and provides an almost instantaneous price for European vanilla options, also the

4 degrees of freedom of the model permit to capture many forms of the volatility smile. However there are some restrictions one should consider. In the next section we explore these restrictions while going through some elements of the proof.

Remark. *Hagan's implied volatility has a small time T asymptotic form and can be written as:*

$$\sigma_B(K,T) = \sigma_B^0(K)(1+\sigma_B^1(K)T) + \mathcal{O}(T^2)$$

3.2 Proof structure

In order to derive equation (3.3), the following general dynamics are considered:

$$\begin{aligned} dF_t &= \widehat{\sigma}C(F_t)dW_t{}^1, \quad F(0) = f \\ d\widehat{\sigma}_t &= \widehat{\nu}\widehat{\sigma}_t dW_t{}^2, \quad \widehat{\sigma}(0) = \alpha \end{aligned} \tag{3.7}$$

where the SABR model dynamics correspond to the special case $C(f) = f^\beta$. The analysis is based on a small volatility expansion, where both the volatility $\widehat{\sigma}$ and the vol of vol $\widehat{\nu}$ are assumed to be small, and therefore can be rewritten as $\widehat{\sigma} \to \epsilon\sigma$, $\widehat{\nu} \to \epsilon\nu$ for $\epsilon > 0$. Thus the dynamics 3.7 become:

$$\begin{aligned} dF_t &= \epsilon\sigma C(F_t)dW_t{}^1, \quad F(0) = f \\ d\sigma_t &= \epsilon\nu\sigma_t dW_t{}^2, \quad \sigma(0) = \alpha \end{aligned} \tag{3.8}$$

The final result is then derived by following the steps below:

1. Derive the partial differential equation satisfied by the correspond transition density function.
2. Solve the PDE using singular perturbation techniques. The option price for the general model 3.8 is then obtained by a applying a simple integration with respect to time.
3. Compute the option price $V(t,f,\sigma_N,K)$ with F following the normal model by setting:

   $$C(f) = 1, \nu = 0 \text{ and } \epsilon\alpha = \sigma_N$$

   Then the normal implied volatility $\sigma_N(K, C(f), \alpha, \nu, \rho)$ is obtained by equating the prices:

   $$V(t,f,\sigma_N,K) = V(t, C(f), \alpha, \nu, \rho)$$

4. We then use this normal implied volatility under Log-normal:

   $$C(f) = f, \nu = 0 \text{ and } \epsilon\alpha = \sigma_B$$

   we write it as $\sigma_N^B(K,f,\sigma_B)$.

   By uniqueness of the price under the general SABR model, and uniqueness of the corresponding normal implied volatility, the Black's implied volatility is obtained by solving:

   $$\sigma_N(K, C(f), \alpha, \nu, \rho) = \sigma_N^B(K, f, \sigma_B)$$

5. Take $C(f) = f^\beta$ in the Balck's implied volatility, thus yielding Hagan's formula.

3.3 Elements of the proof

Following the plan in the previous section, we present some elements of the proof. These elements will allow us to get an understanding of the approximations made, and therefore the limitations on using the final formula.The purpose here is not to derive the intermediate results, but rather present them with some remarks regarding the approximations made.

Step 1

As this step is the starting point, we give it in more detail. We derive the price and the PDE that is to be solved. Suppose the dynamics of the forward swap rate F are given by 3.7, and that the state at time t: $F(t) = f$, $\sigma(t) = \alpha$. Define the transition density function p as:

$$p(t, f, \alpha; T, \bar{F}, \bar{\sigma})d\bar{F}d\bar{\sigma} = P(\bar{F} < F(T) < \bar{F} + d\bar{F}, \bar{\sigma} < \sigma(T) < \bar{\sigma} + d\bar{\sigma} \mid F(t) = f, \sigma(t) = \alpha)$$

$\bar{F}$ and $\bar{\sigma}$ are introduced to distinguish the variables of the transition density function from the processes. Also for simplicity ∂_X denotes the partial derivative operator with respect to X.

The transition density function p satisfies the forward Kolmogorov equation:

$$\partial_T p = \frac{1}{2}\epsilon^2\bar{\sigma}^2\partial_{\bar{F}\bar{F}}p[C^2(\bar{F})p] + \epsilon^2\rho\nu\partial_{\bar{F}\bar{\sigma}}[\bar{\sigma}^2C(\bar{F})p] + \frac{1}{2}\epsilon^2\nu^2\partial_{\bar{\sigma}\bar{\sigma}}[\bar{\sigma}^2 p] \tag{3.9}$$

For $t < T$, and $p = \delta(\bar{F} - f)\delta(\bar{\sigma} - \alpha)$ at $t = T$.

Let $V(t, f, \alpha)$ be the value of a European call option at date t, with $F(t) = f$, $\sigma(t) = \alpha$. Let T_{ex} be the option's exercise date, and let K be the strike. Omitting the discount factor $D(T_{ex})$, the call option value is given by:

$$V(t, f, \alpha) = E\{[F(T_{ex}) - K]^+ \mid F(t) = f, \sigma(t) = \alpha\}$$

By rewriting the transition density function p as:

$$p(t, f, \alpha; T, \bar{F}, \bar{\sigma}) = \delta(\bar{F} - f)\delta(\bar{\sigma} - \alpha) + \int_t^{T_ex} \partial_T p(t, f, \alpha; T, \bar{F}, \bar{\sigma})dT,$$

we can rewrite the option value as:

$$V(t, f, \alpha) = [f - K]^+ + \int_t^{T_ex}\int_K^{\infty}\int_{-\infty}^{\infty}(\bar{F} - K)\partial_T p(t, f, \alpha; T, \bar{F}, \bar{\sigma})d\bar{\sigma}d\bar{F}dT \tag{3.10}$$

Using the equation 3.9, we substitute $\partial_T p$ into 3.10.

$$\begin{aligned} V(t, f, \alpha) = [f - K]^+ + \frac{1}{2}\epsilon^2\int_t^{T_{ex}}\int_K^{\infty}\int_{-\infty}^{\infty}(\bar{F} - K)\bar{\sigma}^2\partial_{\bar{F}\bar{F}}p[C^2(\bar{F})p]d\bar{\sigma}d\bar{F}dT \\ + \epsilon^2\int_t^{T_{ex}}\int_K^{\infty}(\bar{F} - K)\left\{\int_{-\infty}^{\infty}\left(\rho\nu\partial_{\bar{F}\bar{\sigma}}[\bar{\sigma}^2C(\bar{F})p] + \frac{1}{2}\nu^2\partial_{\bar{\sigma}\bar{\sigma}}[\bar{\sigma}^2p]\right)d\bar{\sigma}\right\}d\bar{F}dT \end{aligned}$$

since

$$\partial_{\bar{F}}[\bar{\sigma}^2C(\bar{F})p]\ ,\ \partial_{\bar{\sigma}}[\bar{\sigma}^2p] \to 0 \text{ when } \bar{\sigma} \to \pm\infty$$

Therefore the option price reduces to:

$$V(t, f, \alpha) = [f - K]^+ + \frac{1}{2}\epsilon^2\int_t^{T_{ex}}\int_{-\infty}^{\infty}\int_K^{\infty}\bar{\sigma}^2(\bar{F} - K)\partial_{\bar{F}\bar{F}}[C^2(\bar{F})p]d\bar{F}d\bar{\sigma}dT \tag{3.11}$$

In the expression above the integration order was changed. We perform an integration by parts twice with respect to $\bar{F}$:

$$\begin{aligned}\int_K^\infty \bar{\sigma}^2(\bar{F}-K)\partial_{\bar{F}\bar{F}}[C^2(\bar{F})p]d\bar{F} &= \left[\bar{\sigma}^2(\bar{F}-K)\partial_{\bar{F}}[C^2(\bar{F})p]\right]_K^\infty - \int_K^\infty \bar{\sigma}^2\partial_{\bar{F}}[C^2(\bar{F})p]d\bar{F} \\ &= -\bar{\sigma}^2\left[C^2(\bar{F})p\right]_K^\infty \\ &= \bar{\sigma}^2C^2(K)p\end{aligned}$$

Substituting this expression in 3.11 yields:

$$V(t,f,\alpha) = [f-K]^+ + \frac{1}{2}\epsilon^2C^2(K)\int_t^{T_{ex}}\int_{-\infty}^\infty \bar{\sigma}^2 p(t,f,\alpha,T,K,\bar{\sigma})d\bar{\sigma}dT$$

This can be further simplified by defining:

$$P(t,f,\alpha,T,K) = \int_{-\infty}^\infty \bar{\sigma}^2 p(t,f,\alpha,T,K,\bar{\sigma})d\bar{\sigma}dT$$

Thus the price can the be written as follows:

$$V(t,f,\alpha) = [f-K]^+ + \frac{1}{2}\epsilon^2C^2(K)\int_t^{T_{ex}} P(t,f,\alpha,T,K)dT$$

P then satisfies the backward Kolmogorov equation:

$$\partial_t P + \frac{1}{2}\epsilon^2\alpha^2C^2(f)\partial_{ff}P + \epsilon^2\rho\nu\alpha^2C(f)\partial_{f\alpha}P + \frac{1}{2}\epsilon^2\nu^2\alpha^2\partial_{\alpha\alpha}P = 0, \qquad \forall t < T \tag{3.12}$$

And $P = \alpha^2\delta(f-K)$ for $t = T$.

Since t doesn't appear in this equation, P depends only on $T-t$ and not on t and T separately. Hence, we define the following integration variable:

$$\tau = T - t$$

Using this, we can write the price as follows:

$$V(t,f,\alpha) = [f-K]^+ + \frac{1}{2}\epsilon^2C^2(K)\int_t^{\tau_{ex}} P(\tau,f,\alpha,K)d\tau \tag{3.13}$$

where $\tau_{ex} = T_{ex} - t$. And P is solution of:

$$\partial_\tau P = \frac{1}{2}\epsilon^2\alpha^2C^2(f)\partial_{ff}P + \epsilon^2\rho\nu\alpha^2C(f)\partial_{f\alpha}P + \frac{1}{2}\epsilon^2\nu^2\alpha^2\partial_{\alpha\alpha}P, \qquad \forall\tau > 0 \tag{3.14}$$

And $P = \alpha^2\delta(f-K)$ for $\tau = 0$.

Step 2

Equation 3.14 is solved using singular perturbation techniques. The solution is then substituted in 3.13 to obtain the price under the SABR model with the general function $C(f)$:

$$V(t, f, \alpha) = [f - K]^+ + \frac{|f - K|}{4\sqrt{\pi}} \int_{\frac{x^2}{2\tau_{ex}} - \epsilon^2\theta}^{\infty} q^{-\frac{3}{2}} e^{-q} dq \tag{3.15}$$

with

$$\epsilon^2\theta = \log\left(\frac{\epsilon\alpha z}{f - K}\sqrt{B(0)B(\epsilon\alpha z)}\right) + \log\left(\frac{xI^{\frac{1}{2}}(\epsilon\nu z)}{z}\right) + \frac{1}{4}\epsilon^2\rho\nu\alpha b_1 z^2$$

Pricing a call option with this expression is not practical as the resulting formula is too complicated. At this point the price is given to a precision up to $\mathcal{O}(\epsilon^2)$. One should refer to [1] for the definition of the different terms.

Step 3

We consider the normal model:

$$dF_t = \sigma_N dW_t, \quad F(0) = f$$

This is a special case of the SABR model with $C(f) = 1$, $\nu = 0$ and $\epsilon\alpha = \sigma_N$. In this step we need to express the normal volatility σ_N in terms of the SABR parameters. This volatility can then be used as an input under the standard Normal model. We proceed by inverting the equation:

$$V(t, f, \sigma_N, K) = V(t, C(f), \alpha, \nu, \rho)$$

In order to do this, Hagan considered a small time expansion in τ_{ex} and therefore in T_{ex}.

$$\sigma_N(K) = \sigma_N^0(K)\Big(1 + \sigma_N^1(K)T_{ex}\Big) + \mathcal{O}(T_{ex}^2)$$

This approximation could lead to wrong values for large maturities. This approach leads to the following expression:

$$\begin{aligned}\sigma_N(K) = &\frac{\epsilon\alpha(f - K)}{\int_K^f \frac{df'}{C(f')}}.\left(\frac{\xi}{\hat{x}(\xi)}\right).\\ &\left\{1 + \left[\frac{2\gamma_2 - \gamma_1^2}{24}\alpha^2 C^2(f_{av}) + \frac{1}{4}\rho\nu\alpha\gamma_1 C(f_{av}) + \frac{2 - 3\rho^2}{24}\nu^2\right]\epsilon^2\tau_{ex} + ...\right\}\end{aligned} \tag{3.16}$$

Where

$$f_{av} = \sqrt{fK}, \quad \gamma_1 = \frac{C'(f_{av})}{C(f_{av})}, \quad \gamma_2 = \frac{C''(f_{av})}{C(f_{av})}$$

$$\xi = \frac{\nu}{\alpha}\frac{f - K}{C(f_{av})}, \quad \hat{x}(\xi) = \log\left(\frac{\sqrt{1 - 2\rho\xi + \xi^2} - \rho + \xi}{1 - \rho}\right)$$

The price of a European call option with exercise date τ_{ex} struck at K under the SABR model can be computed by using σ_N in the normal model.

Step 4

we consider the Log-normal model:

$$dF_t = \sigma_B F_t dW_t, \quad F(0) = f$$

The normal implied volatility from the previous step is used under the Log-normal model with $C(f) = f, \nu = 0$ and $\epsilon\alpha = \sigma_B$. In order to obtain σ_B, we need to invert the following equation:

$$\sigma_N(K, C(f), \alpha, \nu, \rho) = \sigma_N^B(K, f, \sigma_B)$$

This results in the following expression:

$$\begin{aligned}\sigma_B(K) =& \frac{\alpha \log\left(\frac{f}{K}\right)}{\int_K^f \frac{df'}{C(f')}} . \left(\frac{\xi}{\hat{x}(\xi)}\right). \\ & \left\{1 + \left[\frac{2\gamma_2 - \gamma_1^2 + \frac{1}{f_{av}^2}}{24}\alpha^2 C^2(f_{av}) + \frac{1}{4}\rho\nu\alpha\gamma_1 C(f_{av}) + \frac{2-3\rho^2}{24}\nu^2\right]\epsilon^2\tau_{ex} + ...\right\}\end{aligned} \tag{3.17}$$

Where

$$\xi = \frac{\nu}{\alpha}\frac{f-K}{C(f_{av})}, \quad \hat{x}(\xi) = \log\left(\frac{\sqrt{1-2\rho\xi+\xi^2} - \rho + \xi}{1-\rho}\right)$$

As for the implied normal volatility, σ_B should be inserted in a Black pricer in order to obtain the price of a European call. This result was obtained by using a further approximation, thus better results should be expected when using σ_N.

Step 5

Now that we are equipped with the normal and log-normal implied volatilities for the general function $C(f)$, we only need to use the special case $C(f) = f^\beta$. Doing so for σ_B yields Hagan's formula.

We won't give the calculations, however we should note that this process requires making another approximation. In fact Hagan considered an expansion in $\log(\frac{f}{K})$, this implies K to be not so far from f.

3.4 Hagan's formula refinments

The derivation of Hagan's formula required several assumptions, these can lead to inaccurate results in some cases. In particular α, ν, T_{ex} are assumed to be small and that K not far from the money. In order to improve Hagan's formula some authors proposed new results (J. Obloj [5], L. Paulot [6] and Antonov et al. [7])

Let's present Antonov et al. results as there are similarities with the approach in the Free Boundary that we explore in the second part of this thesis. In that paper [7] a new exact formula for the zero-correlation case was presented as a two-dimensional integration:

$$\begin{aligned}\mathcal{O}(T,K) &= E[(F_T - K)^+] - (F_0 - K)^+ \\ &= \frac{2}{\pi}\sqrt{KF_0}\left\{\int_{s-}^{s+} \frac{\sin(|\nu|\phi(s))}{\sinh s} G(\nu^2 T, s) ds + \sin(|\nu|\pi)\int_{s+}^{\infty} \frac{e^{-|\nu|\psi(s)}}{\sinh s} G(\nu^2 T, s) ds\right\}\end{aligned} \tag{3.18}$$

where

$$
\begin{aligned}
G(t,s) &= 2\sqrt{2}\frac{e^{-\frac{t}{8}}}{t\sqrt{2\pi t}}\int_s^{\infty} u\, e^{-\frac{u^2}{2t}}\sqrt{\cosh\ u - \cosh\ s}\, du,\\
\phi(s) &= 2\arctan\sqrt{\frac{\sinh^2 s - \sinh^2 s_-}{\sinh^2 s_+ - \sinh^2 s}}\\
\psi(s) &= 2\,\text{arctanh}\sqrt{\frac{\sinh^2 s - \sinh^2 s_+}{\sinh^2 s - \sinh^2 s_-}}\\
s_- &= \text{arcsinh}\left(\frac{|q_K - q_0|}{\mu_0}\right) \text{ and } s_+ = \text{arcsinh}\left(\frac{q_K + q_0}{\mu_0}\right)\\
q_K &= \frac{K^{1-\beta}}{1-\beta}, \qquad q_0 = \frac{f^{1-\beta}}{1-\beta}
\end{aligned}
$$

$\mathcal{O}(T,K)$ is the time value, the call price is given by $E[(F_T - K)^+] = \mathcal{O}(T,K) + (F_0 - K)^+$. We note that function G can be efficiently approximated (see Appendix A. 7), thus reducing the formula to a single integration. These aspects are detailed in the next chapters as the same function is used for prices in both the Free boundary and normal SABR.

The general-correlation price is obtained using a model mapping approach. The idea behind this is to find the general-correlation price using the zero-correlation formula with suitable parameters. In other words, for an initial SABR model defined by $(\alpha, \beta, \nu, \rho)$ there exist a zero-correlation mapping SABR model defined by $(\tilde{\alpha}, \tilde{\beta}, \tilde{\nu}, 0)$ such that:

$$
V_{\text{GC}}(T, K, \alpha, \beta, \nu, \rho) \simeq V_{\text{ZC}}(T, K, \tilde{\alpha}, \tilde{\beta}, \tilde{\nu}, 0) \tag{3.19}
$$

In order to do this the authors match the small time expansion of the time value of both models. This however is not sufficient to completely define the effective parameters, thus some additional constraints must be set. The chosen parametrization can be found in the Advanced analytics for the SABR model [7]. This model mapping approach is a general principle and used in the chapter *the Free Boundary SABR* chapter.

Part II

SABR Extensions to Negative Rates

Chapter 4

Shifted SABR

4.1 Introduction

The shifted SABR is the first and simplest extension of the SABR model to the low-interest-rate environment. Despite several short comings it is widely used by practitioners. It inherits the advantages and intuitive parameters of the SABR model.
Under the shifted SABR the forward rate has the following dynamics:

$$dF_t = \sigma_t (F_t + s)^{\beta} dW_t{}^1, \quad F(0) = f \tag{4.1}$$

$$d\sigma_t = \nu \sigma_t dW_t{}^2, \quad \sigma(0) = \alpha \tag{4.2}$$

where $dW_t{}^1 dW_t{}^2 = \rho dt$ and s a positive deterministic shift.

The shift changes the lower boundary from 0 to $-s$, thereby allowing F_t to reach negative levels. The shift can be either calibrated as an additional parameter of the SABR model or fixed prior to calibration, the first option is not adapted as calibrating the shift only influences the skew (already controlled by β) without adding a new degree of freedom.
In the next sections we will derive the 'shifted' Black-Scholes implied volatility as a function of the SABR process analogously to the classical SABR Black-Scholes implied volatility.

4.2 Call price

In this section we will derive the call option price under the shifted SABR model, assuming that the call option price under the original SABR model is known.

Let s be a positive shift defined such that:

- $F_t + s > 0$ for all $0 \leq t \leq T$.
- $K + s > 0$ for all considered strikes.

We define the shifted forward rate by: $X_t = F_t + s$. We then rewrite the value of a call option as follows:

$$\begin{aligned} E[(F_T - K)^+] &= E[((F_T + s) - (K + s))^+] \\ &= E[(X_T - (K + s))^+] \end{aligned} \tag{4.3}$$

Thus the value of call option on F with strike K is equal to the value of a call option on X with strike $K + s$.

Furthermore using 4.1, we can write

$$dF_t = \sigma_t (F_t + s)^{\beta} dW_t{}^1$$

This is the same as

$$d(F_t + s) = \sigma_t (F_t + s)^{\beta} dW_t{}^1$$

We then rewrite it using X

$$dX_t = \sigma_t X_t{}^{\beta} dW_t{}^1$$

Thus X follows a SABR model dynamics with the same parameters as F:

$$\begin{aligned} dX_t &= \sigma_t X_t{}^{\beta} dW_t{}^1, \quad X(0) = x \\ d\sigma_t &= \nu \sigma_t dW_t{}^2, \quad \sigma(0) = \alpha \end{aligned}$$

where $dW_t{}^1 dW_t{}^2 = \rho dt$.

We conclude that the call option price of a shifted SABR model with the forward f and strike K, is equal the corresponding SABR model (same parameters) call option price with the forward $f + s$ and strike $K + s$.

4.3 Hagan's Approximation for the shifted SABR

In this section we apply the above results to price a call option using Hagan approximation. The forward swap rate F is assumed to follow the shifted SABR dynamics.

Let $BS(f, T, K, \sigma)$ denote the Black-Scholes call option price with forward f, maturity T, strike K ,and volatility σ. The price of a call option on F is then given by:

$$\begin{aligned} E[(F_T - K)^+] &= BS(f + s, T, K + s, \sigma_B(K + s, f + s)) \\ &= (f + s)\mathcal{N}(d_1) - (K + s)\mathcal{N}(d_2) \end{aligned}$$

where the volatility is given by Hagan's approximation, as a function of the initial forward rate, the strike and the shift.

$$\sigma_B(K+s, f+s) = \frac{\alpha}{((f+s)(K+s))^{(1-\beta)/2} \left\{ 1 + \frac{(1-\beta)^2}{24} \log^2\left(\frac{f+s}{K+s}\right) + \frac{(1-\beta)^4}{1920} \log^4\left(\frac{f+s}{K+s}\right) + \dots \right\}} \cdot \left(\frac{z}{x(z)}\right) \cdot \left\{ 1 + \left[\frac{(1-\beta)^2}{24} \frac{\alpha^2}{((f+s)(K+s))^{1-\beta}} + \frac{1}{4} \frac{\rho\beta\nu\alpha}{((f+s)(K+s))^{(1-\beta)/2}} + \frac{2-3\rho^2}{24} \nu^2 \right] T + \dots \right.$$

Here z and $x(z)$ are defined by:

$$\begin{aligned} z &= \frac{\nu}{\alpha} ((f+s)(K+s))^{(1-\beta)/2} \log\left(\frac{f+s}{K+s}\right) \\ x(z) &= \log\left\{ \frac{\sqrt{1 - 2\rho z + z^2} + z - \rho}{1 - \rho} \right\} \end{aligned}$$

4.4 Conclusion

Shifting the SABR model produces excellent results in terms of fitting market data since it inherits the benefits of the original SABR model. Furthermore the results presented in section 4.2 show that it is possible to use any pricing approach under the SABR model only by shifting the forward and the strike. Thus allowing us to benefit from the different methods developed to improve Hagan's initial approximation (*e.g* the Antonov approach).

That being said, the shifted SABR has one major flaw namely fixing the shift prior to calibration. This leads to two direct problems:

- The probability density function of the forward rate degenerates near the lower boundary $-s$ resulting in a delta mass in that region. Therefore while using this model, traders will usually not hedge close to the boundary. They would instead allocate additional funds in reserves in order to pay for their inability to hedge.
- A change in the shift may be required if the rates go lower than expected, this can result in a jump in calibration parameters. Thus jumps in the Greeks of the trades dependent on the swaption or cap volatilities may also occur. To cover for potential losses in such situations, traders are likely to be asked to allocate additional funds in reserves

In the next chapter, we present a new model that doesn't require such adjustments and that seems more naturally adapted to negative interest rates.

Chapter 5

The Free Boundary SABR

5.1 Introduction

A different extension of the SABR model which does not require determining a shift is presented in 'The Free Boundary SABR: Natural Extension to Negative Rates' by Antonov et al. [8]. The forward rate is assumed to have the following dynamics:

$$dF_t = v_t |F_t|^\beta dW_t{}^1, \quad F(0) = F_0 \tag{5.1}$$

$$dv_t = \gamma v_t dW_t{}^2, \quad v(0) = v_0 \tag{5.2}$$

with $0 \leq \beta < \frac{1}{2}$, a free boundary and $dW_t{}^1 dW_t{}^2 = \rho dt$.
The results presented in this section only apply to the case $F_0 > 0$. For $F_0 < 0$, we note that $\tilde{F}_t = -F_t$ satisfies the SDE:

$$\begin{aligned} d\tilde{F}_t &= -dF_t = -|\tilde{F}_t|^\beta v_t dW_1 = |\tilde{F}_t|^\beta v_t d\tilde{W}_1, \\ dv_t &= \gamma v_t dW_2 \end{aligned}$$

where $\tilde{W}_1 = -W_1$. Thus $\tilde{F}_t$ is a free boundary SABR process with parameters $(-F_0, \alpha, \beta, -\rho, \gamma)$. Furthermore as presented in the original paper, the time value $\mathcal{O}^F(T, K)$ (defined below) verifies the following relation

$$\begin{aligned} \mathcal{O}^F(T, K) &= E[(F_T - K)^+] - (F_0 - K)^+ = E[(-K - (-F_T))^+] - (-K - (-F_0))^+ \\ &= E[(\tilde{K} - \tilde{F}_t)^+] - (\tilde{K} - \tilde{F}_0)^+ = \mathcal{O}^{\tilde{F}}(T, \tilde{K}) \end{aligned}$$

where $\tilde{K} = -K$.

This model has the advantage of not bounding how negative rates can become, thus providing a flexible structure for calibration to market data. In the next sections, we present the zero-correlation exact formula for the value of a call option. Next we present the general-correlation case where an approximation is derived using a mimicking model similar to the one used in [7] for the SABR model. Finally we discuss the advantages and drawbacks of this model focusing on calibration to the swaption volatility cube .

5.2 Zero-correlation case

We start by deriving the time value for a CEV (Constant elasticity of variance) model with a free boundary condition. The forward rate then follows the dynamics:

$$dF_t = |F_t|^\beta dW_t, \quad F(0) = F_0 \tag{5.3}$$

Where $0 \leq \beta < \frac{1}{2}$. Such dynamics permit F_T to reach the negative values.
The corresponding forward Kolmogorov equation is:

$$p(t, f) = \frac{1}{2}(|f|^{2\beta} p(t, f))_{ff} \tag{5.4}$$

A solution to this equation with the initial condition $p(0, f) = \delta(f - f_0)$ can be expressed using the reflecting and absorbing solutions for the standard CEV Forward Kolmogorov equation:

$$p(t, f) = \frac{1}{2}(p_R(t, |f|) + \text{sign}(f)p_A(t, |f|) \tag{5.5}$$

where p_R and p_A are respectively the solutions of the forward Kolmogorov equation from the SABR model 5.6 with an respectively an absorbing and reflecting boundary condition.

$$p(t, f) = \frac{1}{f}(f^{2\beta}p(t, f))_{ff} \tag{5.6}$$

We integrate 5.4 with a payoff $h(f)$ over :

$$\partial_t\left(\int_{-\infty}^{+\infty} h(f)p(t, f)\, df\right) = \partial_t\left(\int_{-\infty}^{0} h(f)p(t, f)\, df\right) + \partial_t\left(\int_{0}^{+\infty} h(f)p(t, f)\, df\right)$$

Using the FK equation and integrating by parts twice, we get:

$$\begin{aligned}
\partial_t\left(\int_{0}^{+\infty} h(f)p(t, f)\, df\right) &= \frac{1}{2}\int_{0}^{+\infty} h(f)(f^{2\beta}p(t, f))_{ff}\, df \\
&= \frac{1}{2}[h(f)(f^{2\beta}p(t, f))_f]_0^{+\infty} - \frac{1}{2}\int_{0}^{+\infty} h^{'}(f)(f^{2\beta}p(t, f))_f\, df \\
&= \frac{1}{2}[h(f)(f^{2\beta}p(t, f))_f - h^{'}(f)f^{2\beta}p(t, f)]_0^{+\infty} + \frac{1}{2}\int_{0}^{+\infty} h^{''}(f)f^{2\beta}p(t, f)\, df \\
&= \frac{1}{2}\int_{0}^{+\infty} h^{''}(f)f^{2\beta}p(t, f)\, df + \frac{1}{2}[h^{'}(f)f^{2\beta}p(t, f) - h(f)(f^{2\beta}p(t, f))_f]_{f=0}
\end{aligned}$$

The integral from $-\infty$ to 0 is obtained similarly, thus:

$$\partial_t\left(\int_{-\infty}^{+\infty} h(f)p(t, f)\, df\right) = \frac{1}{2}\int_{-\infty}^{+\infty} h^{''}(f)|f|^{2\beta}p(t, f)\, df + \frac{1}{2}[h^{'}(f)|f|^{2\beta}p(t, f) - h(f)(|f|^{2\beta}p(t, f))_f]_{0-}^{0+}$$

Let us now consider the payoff $h(f) = (f - K)^+$ of a call option, with the left derivative $h^{'}(f) = (f > K)$, and the second derivative $h^{''}(f) = \delta(f - K)$. Integrating between $t = 0$ and $t = T$ yields the option time-value:

$$\begin{aligned}
\mathcal{O}_F(T, K) &= \frac{1}{2}|K|^{2\beta}\int_0^T p(t, K)dt \\
&= \frac{1}{2}|K|^{2\beta}\int_0^T \frac{1}{2}(p_R(t, |K|) + \text{sign}(K)p_A(t, |K|))dt \\
&= \frac{1}{2}(\mathcal{O}_R(t, |K|) + \text{sign}(K)\mathcal{O}_A(T, |K|)).
\end{aligned}$$

The call option time-value for the absorbing case was derived in [9], the reflecting case is given in [8] resulting in the free time-value:

$$\mathcal{O}_F(T,K) = \frac{\sqrt{|K|F_0}}{\pi}\Bigg({}_{K\geq 0}\int_0^{\pi} \frac{\sin(|\nu|\phi)\sin\ \phi}{b-\cos\ \phi} e^{-\frac{\bar{q}(b-\cos\ \phi)}{T}} d\phi$$
$$+ \sin(|\nu|\pi)\int_0^{+\infty} \frac{({}_{K\geq 0}\cosh(|\nu|\psi) + {}_{K<0}\sinh(|\nu|\psi))\sinh\ \psi}{b+\cosh\ \psi} e^{-\frac{\bar{q}(b+\cosh\ \psi)}{T}} d\psi\Bigg) \tag{5.7}$$

Where $\bar{q} = \frac{|F_0K|^{1-\beta}}{(1-\beta)^2}$, $\nu = -\frac{1}{2(1-\beta)}$ and $b = \frac{|F_0|^{2(1-\beta)}+|K|^{2(1-\beta)}}{2|F_0K|^{1-\beta}}$.

Equipped with this result we can derive the exact option-time value of the zero-correlation SABR using that:

$$\mathcal{O}_F^{SABR}(T,K) = E[\mathcal{O}_F^{CEV}(\tau_T,K)], \tag{5.8}$$

where the stochastic time "$\tau_T = \int_0^T v_t^2 dt$" is the cumulative variance for the geometric Brownian motion v_t. Substituting 5.7 into this equation yields:

$$\mathcal{O}_F^{SABR}(T,K) = E[\mathcal{O}_F^{CEV}(\tau_T,K)]$$
$$= \frac{\sqrt{|K|F_0}}{\pi}\Bigg({}_{K\geq 0}\int_0^{\pi} \frac{\sin(|\nu|\phi)\sin\ \phi}{b-\cos\ \phi} [e^{-\frac{\bar{q}(b-\cos\ \phi)}{\tau_T}}] d\phi$$
$$+ \sin(|\nu|\pi)\int_0^{+\infty} \frac{({}_{K\geq 0}\cosh(|\nu|\psi) + {}_{K<0}\sinh(|\nu|\psi))\sinh\ \psi}{b+\cosh\ \psi} [e^{-\frac{\bar{q}(b+\cosh\ \psi)}{\tau_T}}] d\psi\Bigg) \tag{5.9}$$

Here we can see that we need to compute an expectation over the stochastic time of the form $E[e^{-\frac{\lambda}{\tau}}]$. To accomplish this, we use the moment generating function of the inverse stochastic time derived in [9].

$$E[e^{-\frac{\lambda}{\tau}}] = \frac{G(T\gamma^2, s)}{\cosh\ s}, \tag{5.10}$$

$$G(t,s) = 2\sqrt{2}\frac{e^{-\frac{t}{8}}}{t\sqrt{2\pi t}}\int_s^{\infty} u\, e^{-\frac{u^2}{2t}}\sqrt{\cosh\ u - \cosh\ s}\, du, \tag{5.11}$$

$$s(\lambda) = \sinh^{-1}\left(\frac{\sqrt{2\lambda}\gamma}{v_0}\right) \tag{5.12}$$

Applying this result to equation 5.9 with $\lambda_1 = \bar{q}(b-\cos\ \phi)$ and $\lambda_2 = \bar{q}(b+\cosh\ \psi)$ yields:

$$\mathcal{O}_F^{SABR}(T,K) = \frac{1}{2}\sqrt{|KF_0|}\{_{K\geq 0}A_1 + \sin(|\nu|\pi)A_2\} \tag{5.13}$$

With the integrals A_1 and A_2:

$$A_1 = \int_0^{\pi} \frac{\sin\ \phi\ \sin(|\nu|\phi)}{b - \cos\ \phi} \frac{G(T\gamma^2, s(\phi))}{\cosh\ s(\phi)} d\phi, \tag{5.14}$$

$$A_2 = \int_0^{\infty} \frac{\sinh\ \psi\ ({}_{K\geq 0}\cosh(|\nu|\psi) + {}_{K<0}\sinh(|\nu|\psi))}{b + \cosh\ \psi} \frac{G(T\gamma^2, s(\psi))}{\cosh\ s(\psi)} d\psi \tag{5.15}$$

And the parametrization:

$$\begin{aligned}
\overline{q} &= \frac{|F_0K|^{1-\beta}}{(1-\beta)^2}, \\
\nu &= -\frac{1}{2(1-\beta)}, \\
b &= \frac{|F_0|^{2(1-\beta)} + |K|^{2(1-\beta)}}{2|F_0K|^{1-\beta}}, \\
\sinh\ s(\phi) &= \frac{\gamma}{v_0}\sqrt{2\overline{q}(b - \cos\phi)}, \\
\sinh\ s(\psi) &= \frac{\gamma}{v_0}\sqrt{2\overline{q}(b + \cosh\psi)}
\end{aligned}$$

This exact pricing formula contains double integrals leading to time consuming computations, as a result making calibration impractical. Fortunately it turns out that the function G can be well approximated by a closed formula, thus reducing the problem to simple integrals.

5.3 General-correlation case

Antonov shows that the zero and general correlation free boundary SABR models have the same small time expansion for the option value, thus in order to obtain the price in the general correlation case we use a model mapping approach.

To do so we define a zero-correlation free SABR (mimicking) model:

$$\begin{aligned}
d\tilde{F}_t &= |\tilde{F}_t|^{\tilde{\beta}}\tilde{v}_t d\tilde{W}_1, \\
d\tilde{v}_t &= \tilde{\gamma}\tilde{v}_t d\tilde{W}_2
\end{aligned}$$

where $E[d\tilde{W}_1 d\tilde{W}_2] = 0$. The parameters of this model are set such that,

$$V_{\text{call}}^{\text{ZC}}(T, K, \tilde{v}_0, \tilde{\beta}, \tilde{\gamma}, 0) \simeq V_{\text{call}}^{\text{GC}}(T, K, v_0, \beta, \gamma, \rho)$$

The zero-correlation effective parameters $(\tilde{v}_0, \tilde{\beta}, \tilde{\gamma})$ are calculated by matching the small time asymptotics with non-zero correlation model. This approach allows a projection of the general correlation into the zero correlation, for which there is an exact pricing formula. Here we use the effective parameters given in [8], and initially derived for the SABR model in [7].

- Strike-independent parameters: $\tilde{\beta}$, $\tilde{\gamma}$

$$\tilde{\beta} = \beta, \tag{5.16}$$

$$\tilde{\gamma}^2 = \gamma^2 - \frac{3}{2}\{\gamma^2\rho^2 + v_0\gamma\rho(1-\beta)F_0^{\beta-1}\} \tag{5.17}$$

- Strike-dependent parameters: $\tilde{v}_0$

  The initial stochastic volatility is calculated as a time expansion:

$$\tilde{v}_0 = \tilde{v}_0^{(0)} + T\tilde{v}_0^{(1)} + ... \tag{5.18}$$

  The leading volatility term can be expressed as

$$\tilde{v}_0^{(0)} = \frac{2\Phi\, \delta\tilde{q}\, \tilde{\gamma}}{\Phi^2 - 1} \tag{5.19}$$

  where

$$\Phi = \left(\frac{v_{min} + \rho v_0 + \gamma\delta q}{(1+\rho)v_0}\right)^{\frac{\tilde{\gamma}}{\gamma}} \text{ and } v_{min}^2 = \gamma^2\delta q^2 + 2\gamma\rho\delta q v_0 + v_0^2$$

  for

$$\delta q = \frac{|k|^{1-\beta} - |F_0|^{1-\beta}}{1-\beta} \text{ and } \delta\tilde{q} = \frac{|k|^{1-\tilde{\beta}} - |F_0|^{1-\tilde{\beta}}}{1-\tilde{\beta}}$$

  where k is set to:

$$k = \max(K, 0.1F_0)$$

  The first-order correction is more complicated and is given by:

$$\frac{\tilde{v}_0^{(1)}}{\tilde{v}_0^{(0)}} = \tilde{\gamma}^2\sqrt{1+\tilde{R}^2}\frac{\frac{1}{2}\ln\left(\frac{v_0 v_{min}}{\tilde{v}_0^{(0)}\tilde{v}_{min}}\right) - \mathcal{B}_{min}}{\tilde{R}\ln\left(\sqrt{1+\tilde{R}^2}+\tilde{R}\right)} \text{ for } \tilde{R} = \frac{\delta q\tilde{\gamma}}{\tilde{v}_0^{(0)}},$$

  where

$$\tilde{v}_{min} = \sqrt{\tilde{\gamma}^2\delta q^2 + ({\tilde{v}_0}^{(0)})^2}$$

  and $\mathcal{B}_{min}$ is the so-called parallel transport, defined as

$$\mathcal{B}_{min} = -\frac{\beta}{2(1-\beta)}\frac{\rho}{\sqrt{1-\rho^2}}(\pi - \varphi_0 - \arccos\rho - I)$$

  Here we have

$$L = \frac{v_{min}}{q\gamma\sqrt{1-\rho^2}} \text{ and } \varphi_0 = \arccos\left(-\frac{\delta q\gamma + v_0\rho}{v_{min}}\right)$$

  and

$$I = \begin{cases} \frac{2}{\sqrt{1-L^2}}\left(\arctan\left(\frac{u_0+L}{\sqrt{1-L^2}}\right) - \arctan\left(\frac{L}{\sqrt{1-L^2}}\right)\right) & \text{for } L < 1 \\ \frac{1}{\sqrt{L^2-1}}\ \ln\left(\frac{u_0(L+\sqrt{L^2-1})+1}{u_0(L-\sqrt{L^2-1})+1}\right) & \text{for } L > 1 \end{cases} \tag{5.20}$$

As mentioned above, this approach is based on matching parameters of the zero and general correlation free boundary SABR models by using a small time expansion. This yields an unstable parametrization, as it is possible to find a set of parameters that results in a negative effective initial volatility $\tilde{v}_0$. This is more likely to occur for long dated options, as the small-time hypothesis will no longer be valid. In such cases, the contribution of first-order correction becomes more important than that of the leading volatility term.

Due to such constraints we choose not calibrate this model to market data, but instead use the Mixture SABR model (6). However, in the remainder of this chapter, we address the ATM (At-The-Money) limits $(K \to F_0)$ of this parametrization, and perform numerical experiments that highlights some interesting aspects of the Free Boundary that motivates the study in the next chapter.

ATM limit

We note that the initial stochastic $\tilde{v}_0$ parametrization requires taking careful limits ATM. These limits are not provided in the original paper, since these results are very similar to what was obtained in [7] for the classical SABR model. Here we use the same approach as in Appendix C of that article.

Proposition:

The ATM limits for the initial volatility are given by the following results:

• The leading order ATM limit:

$$\tilde{v}_0^{(0)}\big|_{K=F_0} = v_0 \tag{5.21}$$

• The first correction ATM limit:

$$\frac{\tilde{v}_0^{(1)}}{\tilde{v}_0^{(0)}}\bigg|_{K=F_0} = \frac{1}{12}\left(1 - \frac{\tilde{\gamma}^2}{\gamma^2} - \frac{3}{2}\rho^2\right)\gamma^2 + \frac{1}{4}\beta\rho v_0 \gamma F_0^{\beta-1} \tag{5.22}$$

Proof. In the following we assume that $\tilde{\beta} = \beta$.
Since the forward rate F_0 is assumed to be positive we have $k|_{K=F_0} = F_0$. which yields:

$$\delta q = \delta\tilde{q} = \frac{|k|^{1-\beta} - |F_0|^{1-\beta}}{1-\beta} \to 0 \quad \text{as} \quad K \to F_0$$

Let us expand the different expressions defined in 24. We define $\delta z = \gamma\frac{\delta q}{v_0}$, and consider the small δz expansions:

$$\begin{aligned}
\bullet\ v_{min} &= \sqrt{\gamma^2\delta q^2 + 2\gamma\rho\delta q v_0 + v_0^2} \\
&= v_0\sqrt{1 + 2\rho\delta z + \delta z^2} \\
&= v_0\big(1 + \rho\delta z + \frac{1}{2}(1-\rho^2)\delta z^2 - \frac{1}{2}\rho(1-\rho^2)\delta z^3 + O(\delta z^4)\big)
\end{aligned}$$

$$\begin{aligned}
\bullet\ \Phi &= \left(\frac{v_{min} + \rho v_0 + \gamma\delta q}{(1+\rho)v_0}\right)^a \text{ where } a = \frac{\tilde{\gamma}}{\gamma} \\
&= \left(\frac{\delta z + \rho + \frac{v_{min}}{v_0}}{1+\rho}\right)^a \\
&= \left(\frac{1}{1+\rho}\left[\delta z + \rho + 1 + \rho\delta z + \frac{1}{2}(1-\rho^2)\delta z^2 - \frac{1}{2}\rho(1-\rho^2)\delta z^3 + O(\delta z^4)\right]\right)^a \\
&= \left(\frac{1}{1+\rho}\left[(1+\rho) + (1+\rho)\delta z + \frac{1}{2}(1-\rho^2)\delta z^2 - \frac{1}{2}\rho(1-\rho^2)\delta z^3 + O(\delta z^4)\right]\right)^a \\
&= \big(1 + \delta z + \frac{1}{2}(1-\rho)\delta z^2 - \frac{1}{2}\rho(1-\rho)\delta z^3 + O(\delta z^4)\big)^a \\
&= 1 + a\delta z + \frac{a}{2}(1-\rho)\delta z^2 - \frac{a}{2}\rho(\rho-1)\delta z^3 + \frac{a(a-1)}{2}\delta z^2 + \frac{a(a-1)(1-\rho)}{2}\delta z^3 \\
&+ \frac{a(a-1)(a-2)}{6}\delta z^3 + O(\delta z^4) \\
&= 1 + a\delta z + \frac{a(a-\rho)}{2}\delta z^2 + \frac{a}{6}\big(3\rho(1-\rho) + 3(a-1)(1-\rho) + (a-1)(a-2)\big)\delta z^3 + O(\delta z^4) \\
&= 1 + a\delta z + \frac{1}{2}a(a-\rho)\delta z^2 + \frac{a}{6}(3\rho^2 - 1 - 3a\rho + a^2)\delta z^3 + O(\delta z^4)
\end{aligned}$$

Using this result, we can compute the expansion of $\Phi^2 - 1$:

$$\Phi^2 = 1 + 2a\delta z + a(a-\rho)\delta z^2 + \frac{a}{3}(3\rho^2 - 1 - 3a\rho + a^2)\delta z^3 + a^2\delta z^2 + a^2(a-\rho)\delta z^3 + O(\delta z^4)$$
$$\Phi^2 - 1 = 2a\delta z + a(a-\rho)\delta z^2 + \frac{a}{3}(3\rho^2 - 1 - 3a\rho + a^2 + 3a^2 - 3\rho a)\delta z^3 + O(\delta z^4)$$
$$\Phi^2 - 1 = 2a\delta z\Big[1 + (a - \frac{\rho}{2})\delta z + \frac{1}{6}(3\rho^2 - 1 - 6a\rho + 4a^2)\delta z^2 + O(\delta z^3)\Big]$$
$$\frac{2a\delta z}{\Phi^2 - 1} = 1 - (a - \frac{\rho}{2})\delta z - \frac{1}{6}(3\rho^2 - 1 - 6\rho a + 4a^2)\delta z^2 + (a^2 - \rho a + \frac{\rho^2}{4})\delta z^2 + O(\delta z^3)$$
$$\frac{2a\delta z}{\Phi^2 - 1} = 1 + (\frac{\rho}{2} - a)\delta z + \frac{1}{6}(1 + 2a^2 - \frac{3}{2}\rho^2)\delta z^2 + O(\delta z^3)$$

The expansion of the leading volatility term $\tilde{v}_0^{(0)}$ is then given by:

$$\bullet\ \tilde{v}_0^{(0)} = \frac{2\Phi\delta q\tilde{\gamma}}{\Phi^2 - 1}$$
$$= v_0\Phi\frac{2a\delta z}{\Phi^2 - 1}$$
$$= v_0\Big(1 + a\delta z + \frac{a}{2}(a-\rho)\delta z^2 + O(\delta z^3)\Big)\Big(1 + (\frac{\rho}{2} - a)\delta z + \frac{1}{6}(1 + 2a^2 - \frac{3}{2}\rho^2)\delta z^2 + O(\delta z^3)\Big)$$
$$= v_0\Big[1 + \frac{\rho}{2}\delta z + \big(a(\frac{\rho}{2} - a) + \frac{a}{2}(a-\rho) + \frac{1}{6}(1 + 2a^2 - \frac{3}{2}\rho^2)\big)\delta z^2 + O(\delta z^3)\Big]$$
$$= v_0\Big[1 + \frac{\rho}{2}\delta z + \frac{1}{6}(1 - a^2 - \frac{3}{2}\rho^2)\delta z^2 + O(\delta z^3)\Big]$$

Particularly the leading volatility ATM limit is:

$$\tilde{v}_0^{(0)}\big|_{K=F_0} = v_0 \tag{5.23}$$

Let's now expand the correction term:

$$\frac{\tilde{v}_0^{(1)}}{\tilde{v}_0^{(0)}} = \tilde{\gamma}^2\sqrt{1+\tilde{R}^2}\frac{\frac{1}{2}\ln\Big(\frac{v_0 v_{min}}{\tilde{v}_0^{(0)}\tilde{v}_{min}}\Big) - \mathcal{B}_{min}}{\tilde{R}\ \ln\big(\sqrt{1+\tilde{R}^2} + \tilde{R}\big)} \quad \text{for } \tilde{R} = \frac{\delta q\tilde{\gamma}}{\tilde{v}_0^{(0)}},$$

First, we rewrite $\tilde{R}$ in terms of Φ:

$$\tilde{R} = \frac{\delta q\tilde{\gamma}}{\tilde{v}_0^{(0)}} = \frac{\delta q\tilde{\gamma}}{2\Phi\delta q\tilde{\gamma}}(\Phi^2 - 1) = \frac{\Phi^2 - 1}{2\Phi}$$

Let N denote the numerator and D the denominator of the correction term. We have:

$$D = \frac{\sqrt{1+\tilde{R}^2}}{\tilde{R}}\ln\big(\sqrt{1+\tilde{R}^2} + \tilde{R}\big)$$

where

$$
\begin{aligned}
\sqrt{1+\tilde{R}^2} &= \sqrt{1+\frac{(\Phi^2-1)^2}{4\Phi^2}} \\
&= \frac{1}{2\Phi}\sqrt{4\Phi^2+(\Phi^2-1)^2} \\
&= \frac{1}{2\Phi}\sqrt{(\Phi^2+1)^2} \\
&= \frac{\Phi^2+1}{2\Phi}
\end{aligned}
$$

Using this expression yields the following :

$$
\begin{aligned}
D &= \frac{\Phi^2+1}{2\Phi}\frac{2\Phi}{\Phi^2-1} = \frac{\Phi^2+1}{\Phi^2-1}\ln\big(\frac{\Phi^2+1}{2\Phi}+\frac{\Phi^2-1}{2\Phi}\big) \\
&= \frac{\Phi^2-1}{\Phi^2+1}\ln(\Phi)
\end{aligned}
$$

Furthermore:

$$\Phi = 1 + a\delta z + O(\delta z^2)$$

which gives us

$$\frac{\Phi^2-1}{\Phi^2+1} = a\delta z + O(\delta z^2) \text{ and } \ln(\Phi) = a\delta z + O(\delta z^2)$$

Thus the denominator has the following expansion:

$$D = a^2\delta z^2 + O(\delta z^3) \tag{5.24}$$

Let's now expand the numerator

$$N = \tilde{\gamma}^2\frac{1}{2}\ \ln\left(\frac{v_0 v_{min}}{\tilde{v}_0^{(0)}\tilde{v}_{min}}\right) - \mathcal{B}_{\min}$$

We have:

$$\ln\left(\frac{v_0 v_{min}}{\tilde{v}_0^{(0)}\tilde{v}_{min}}\right) = \ln(v_0 v_{min}) - \ln(\tilde{v}_0^{(0)}\tilde{v}_{min})$$

- $$
\begin{aligned}
\ln(v_0 v_{min}) &= \ln(v_0^2) + \ln(1+\rho\delta z + \frac{1}{2}(1-\rho^2)\delta z^2 + O(\delta z^3)) \\
&= 2\ln v_0 + \rho\delta z + \frac{1}{2}(1-\rho^2)\delta z^2 - \frac{\rho^2}{2}\delta z^2 + O(\delta z^3) \\
&= 2\ln v_0 + \rho\delta z + \frac{1}{2}(1-2\rho^2)\delta z^2 + O(\delta z^3)
\end{aligned}
$$

The second term is more complex. We start by expressing $\tilde{v}_{\min}$ using terms with know expansions.

$$\tilde{v}_{\min} = \sqrt{\tilde{\gamma}^2\delta q^2 + (\tilde{v}_0^{(0)})^2} \quad = \tilde{v}_0^{(0)}\sqrt{1+\tilde{R}^2} = \tilde{v}_0^{(0)}\frac{\Phi^2+1}{2\Phi}$$

Thus,

$$\ln(\tilde{v}_0^{(0)}\tilde{v}_{min}) = 2\ln\tilde{v}_0^{(0)} + \ln(\Phi^2+1) - \ln(2\Phi)$$

We then use the derived expansions of Φ, Φ^2 and $\tilde{v}_0^{(0)}$.

$$\begin{aligned}\ln(\Phi^2+1) &= \ln 2 + a\delta z + a(a-\frac{\rho}{2})\delta z^2 - \frac{a^2}{2}\delta z^2 + O(\delta z^3)\\ &= \ln 2 + a\delta z + \frac{a}{2}(a-\rho)\delta z^2 + O(\delta z^3)\end{aligned}$$

$$\begin{aligned}\ln(2\Phi) &= \ln 2 + a\delta z + \frac{a}{2}(a-\rho)\delta z^2 - \frac{a^2}{2}\delta z^2 + O(\delta z^3)\\ &= \ln 2 + a\delta z - \frac{a\rho}{2}\delta z^2 + O(\delta z^3)\end{aligned}$$

$$\begin{aligned}2\ln\tilde{v}_0^{(0)} &= 2\ln v_0 + \rho\delta z + \frac{1}{3}(1-a^2-\frac{3}{2}\rho^2)\delta z^2 - \frac{\rho^2}{4}\delta z^2 + O(\delta z^3)\\ &= 2\ln v_0 + \rho\delta z + \frac{1}{3}(1-a^2-\frac{9}{4}\rho^2)\delta z^2 + O(\delta z^3)\end{aligned}$$

Therefore we obtain:

- $$\begin{aligned}\ln(\tilde{v}_0^{(0)}\tilde{v}_{\min}) &= 2\ln\tilde{v}_0^{(0)} + \ln(\Phi^2+1) - \ln(2\Phi)\\ &= 2\ln\tilde{v}_0^{(0)} + \frac{a^2}{2}\delta z^2 + O(\delta z^3)\\ &= 2\ln v_0 + \rho\delta z + \frac{1}{3}(1+\frac{a^2}{2}-\frac{9}{4}\rho^2)\delta z^2 + O(\delta z^3)\end{aligned}$$

Hence

$$\begin{aligned}\ln\left(\frac{v_0 v_{min}}{\tilde{v}_0^{(0)}\tilde{v}_{min}}\right) &= \ln(v_0 v_{min}) - \ln(\tilde{v}_0^{(0)}\tilde{v}_{min})\\ &= \frac{1}{6}(1-a^2-\frac{3}{2}\rho^2)\delta z^2 + O(\delta z^3)\end{aligned}$$

Furthermore, we use the expansion of the parallel transport $\mathcal{B}_{\min}$ derived in [7].

$$\begin{aligned}\mathcal{B}_{\min} &= -\frac{1}{4}\frac{\beta}{1-\beta}\rho\frac{\delta q^2\gamma}{q_0 v_0} + O(\delta q^3) \;\text{ where }\; q_0 = \frac{F_0^{1-\beta}}{1-\beta}\\ &= -\frac{1}{4}\frac{\beta}{1-\beta}\rho\frac{v_0}{q_0\gamma}\delta z^2 + O(\delta z^3)\end{aligned}$$

Thus the nominator has the following expansion:

$$N = \tilde{\gamma}^2\frac{1}{12}(1-a^2-\frac{3}{2}\rho^2)\delta z^2 + \frac{1}{4}\frac{\beta}{1-\beta}\rho\frac{v_0}{q_0\gamma}\delta z^2 + O(\delta z^3) \tag{5.25}$$

The expansions 5.24 and 5.25 lead to the correction term ATM limit:

$$\left.\frac{\tilde{v}_0^{(1)}}{\tilde{v}_0^{(0)}}\right|_{K=F_0} = \frac{1}{a^2\delta z^2}\left[\tilde{\gamma}^2\frac{1}{12}(1-a^2-\frac{3}{2}\rho^2)\delta z^2 + \frac{1}{4}\frac{\beta}{1-\beta}\rho\frac{v_0}{q_0\gamma}\delta z^2\right]$$

Substituting a, q_0 and δz by their respective expressions, yields the ATM correction:

$$\left.\frac{\tilde{v}_0^{(1)}}{\tilde{v}_0^{(0)}}\right|_{K=F_0} = \frac{1}{12}\left(1 - \frac{\tilde{\gamma}^2}{\gamma^2} - \frac{3}{2}\rho^2\right)\gamma^2 + \frac{1}{4}\beta\rho v_0 \gamma F_0^{\beta-1} \tag{5.26}$$

□

5.4 Numerical Experiments

In this section we will present the numerical results obtained using our implementation of the Free Boundary. We will start by presenting the probability density functions for a given set of parameters, then we will compare our call prices results to the ones in the original article [8]. Finally we will focus on the particular case of zero strike and/or zero forward.

5.4.1 Free Boundary SABR PDF

Let $C(K,T) = E[(F_T - K)^+]$ denote the call price with maturity T and strike K. And let f_F denote the implied probability density function. In order to derive it, we use the approach of Breeden and Litzenberger [10]:

$$\begin{aligned} C(K,T) = E[(F_T - K)^+] &= \int_{-\infty}^{+\infty} (x-K)^+ f_F(x)dx \\ &= \int_K^{+\infty} (x-K) f_F(x)dx \end{aligned}$$

Taking the derivative with respect to the strike yields:

$$\begin{aligned} \frac{\partial C(K,T)}{\partial K} &= -Kf_F(K) + Kf_F(K) - \int_K^{+\infty} f_F(x)dx \\ \frac{\partial C(K,T)}{\partial K} &= -\int_K^{+\infty} f_F(x)dx \end{aligned} \tag{5.27}$$

Thus the PDF is obtained using the second derivative with respect to the strike.

$$f_F(K) = \frac{\partial^2 C(K,T)}{\partial K^2}$$

Furthermore we can write the distribution function of F using equation 5.27:

$$P_F(F \leq K) = 1 + \frac{\partial C(K,T)}{\partial K}$$

Since we do not have closed formulas for this expressions under the Free Boundary SABR model, in practice we will use the central difference approximation:

$$P_F(F \leq K) = 1 + \frac{\partial C(K,T)}{\partial K} \approx 1 + \frac{C(K+\delta K,T) - C(K-\delta K,T)}{2\,\delta K} \tag{5.28}$$

$$f_F(K,T) = \frac{\partial^2 C(K,T)}{\partial K^2} \approx \frac{C(K+\delta K,T) - 2C(K,T) + C(K-\delta K,T)}{\delta K^2} \tag{5.29}$$

Parameter	Input I	Input II
F_0	50 bps	50 bps
v_0	0.6 $F_0^{1-\beta}$	0.6 $F_0^{1-\beta}$
γ	0.3	0.3
ρ	-0.3	-0.3
β	**0.1**	**0.25**
T	3Y	3Y

Table 5.1: Implied probability density function inputs

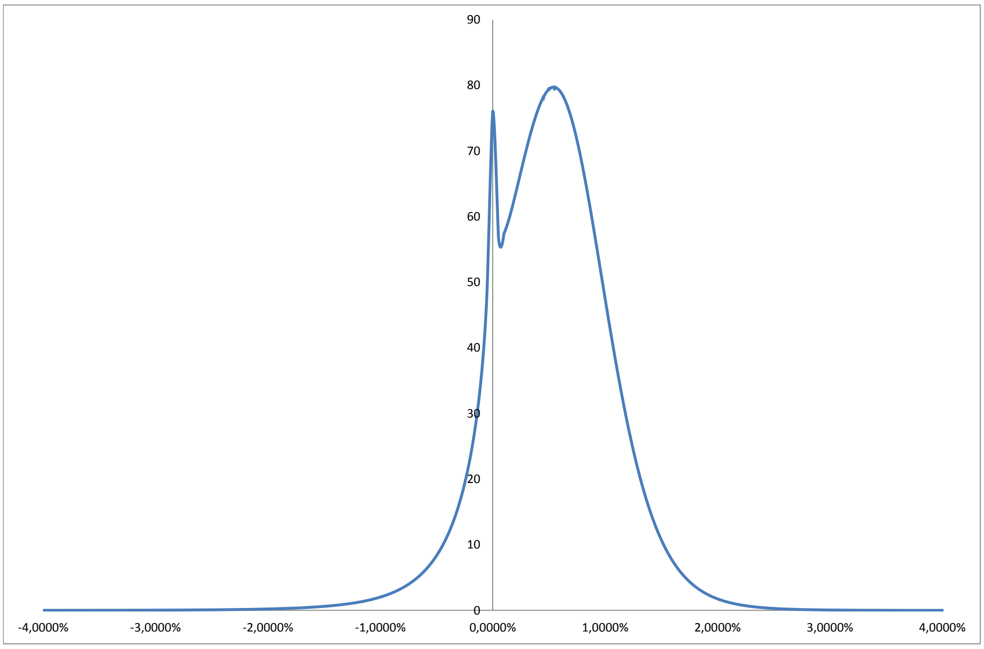


Figure 5.1: Free Boundary SABR implied PDF for Input I: $\beta = 0.1$

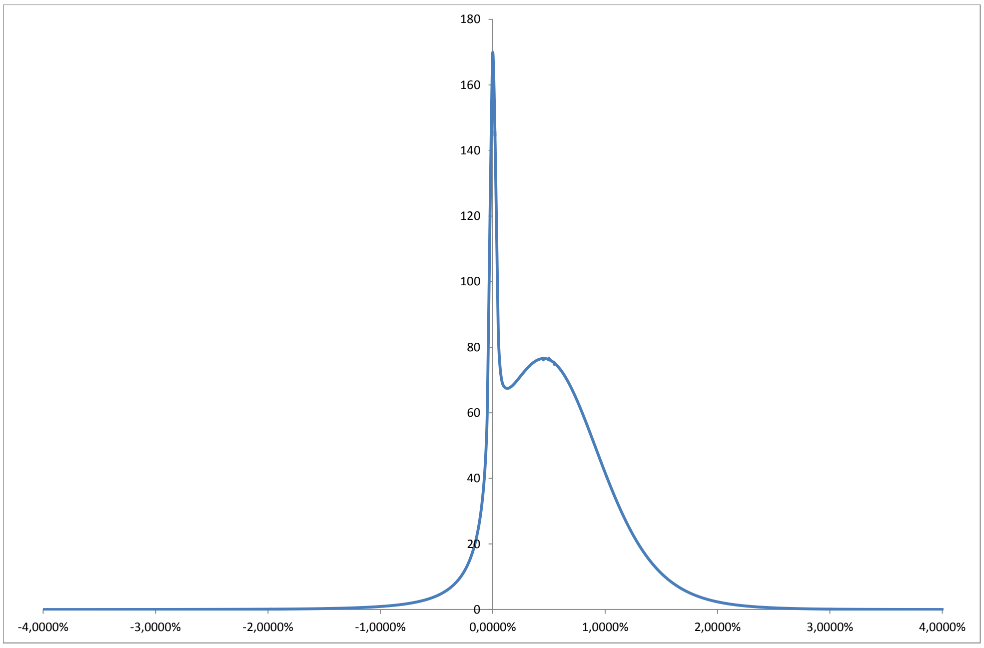


Figure 5.2: Free Boundary SABR implied PDF for Input II: $\beta = 0.25$

In the following, we consider the same inputs used in the original article 5.1.

The Figures 5.1 & 5.2 represent the implied probability density function respectively for Input I and II. From these we can see that a key feature of the Free Boundary model is that the forward rates has a PDF consistent with the natural market perception. The forward rate have a non zero probability of reaching negative territory and tends to stick around 0. We point out that the intensity of this stickiness (spike at 0 of the implied density) is controlled by the parameter β.

Therefore, unlike the original/Shifted SABR model, where the parameter β doesn't change the quality of calibration, we expect it here to have a significant contribution.

Another interesting way to compare this model with shifted SABR is to look at the implied distribution function. If we set for example the shift at 2% then:

$$P_F^{\text{Shifted}}(F \leq -2\%) = 0$$

Whereas,

$$P_F^{\text{Free}}(F \leq -2\%)_{\text{Input I}} = 0.0153$$
$$P_F^{\text{Free}}(F \leq -2\%)_{\text{Input II}} = 0.0202$$

Thus under the Free Boundary dynamics the forward rate isn't bounded by a lower limit like the Shifted SABR. This feature is very important since we cannot be certain that the rates will not go beneath a fixed threshold.

5.4.2 Comparison with Antonov's results

The implementation of Hagan's approximation for the SABR model is straightforward, as it only uses elementary trigonometric functions, whereas the implementation of the Free boundary SABR can produce different results depending on the chosen numerical methods. In this section we compare our results with those presented in [8].

We compute the normal implied volatility in basis points ($1\text{bps} = 0.01\%$) for European call options with the parameters shown in Table 5.2.

Parameter	Symbol	Value for Input I	Value for Input II
Rate Initial Value	F_0	50 bps	1%
SV Initial Value	v_0	0.6 $F_0^{1-\beta}$	0.3 $F_0^{1-\beta}$
Vol-of-Vol	γ	0.3	0.3
Correlations	ρ	-0.3	-0.3
Skews	β	0.1	0.25
Maturities	T	3Y	10Y

Table 5.2: Setups for the free-boundary SABR model

In Table 5.3 & 5.4 , we compare our results with Antonov's analytical results (A) and the Monte Carlo simulations (MC) both from [8]. The normal implied volatilities are obtained by *numerically* inverting the prices.
From these Tables (5.3 & 5.4) we can see that we obtain the same results as in [8] with the highest difference being $0.01\ bps$.

Here we have used the Integral version of function G, but we obtain exactly the same results with the approximation.

5.4.3 Special case: F = 0 & K = 0

In this section we will go through the special case of zero strikes and zero forwards. It was pointed out in [2], that the call price is a smooth function of K and F_0, and that one can show:

$$\begin{aligned} C(T, K, F_0) &\underset{K\to 0}{=} C_1 + C_2 K + C_3 |K|^{2(1-\beta)} + ..., \\ C(T, K, F_0) &\underset{F_0\to 0}{=} C_1^{'} + C_2^{'} F_0 + C_3^{'} |F_0|^{2(1-\beta)} + ..., \end{aligned} \tag{5.30}$$

where the constants C_i and $C_i^{'}$ depend on the model parameters.

We will only address the zero correlation case, consequently we are not concerned by the asymptotics for the general correlation parametrization.

From the zero correlation time value formula 5.13, we see that the cases $K = 0$ and $F_0 = 0$ are problematic. However any other value for the strike or the forward rate no matter how small (in absolute value) can be directly substituted into the pricing formula without any problem.

Assuming the limit exists it must equal the left and right hand limits. We use a heuristic whereby the limit is computed as the average of the estimates of the left and right hand limits. Thus the prices at $K = 0$ and $F = 0$ are obtained by:

$$\begin{aligned} C(T, K = 0, F_0) &= \frac{C(T, \epsilon, F_0) + C(T, -\epsilon, F_0)}{2} \\ C(T, K, F_0 = 0) &= \frac{C(T, K, \epsilon) + C(T, K, -\epsilon)}{2} \end{aligned}$$

$\frac{K}{F_0}$	K	σ^N	σ^N_A	σ^N_{MC}	diff A	diff MC
-0.95	-0.48%	30.87	30.87	30.93	0.00	-0.06
-0.80	-0.40%	29.83	29.83	29.95	0.00	-0.12
-0.65	-0.33%	28.79	28.80	28.97	-0.01	-0.18
-0.50	-0.25%	27.79	27.79	27.99	0.00	-0.20
-0.35	-0.18%	26.82	26.83	27.04	-0.01	-0.22
-0.20	-0.10%	25.95	25.95	26.15	0.00	-0.20
-0.05	-0.03%	25.30	25.30	25.46	0.00	-0.16
0.10	0.05%	25.77	25.77	25.85	0.00	-0.08
0.25	0.13%	26.63	26.63	26.69	0.00	-0.06
0.40	0.20%	27.33	27.33	27.39	0.00	-0.06
0.55	0.28%	27.90	27.90	27.97	0.00	-0.07
0.70	0.35%	28.38	28.38	28.45	0.00	-0.07
0.85	0.43%	28.80	28.80	28.87	0.00	-0.07
1.00	**0.50%**	**29.17**	**29.18**	**29.25**	**-0.01**	**-0.08**
1.15	0.58%	29.53	29.53	29.60	0.00	-0.07
1.30	0.65%	29.87	29.87	29.94	0.00	-0.07
1.45	0.73%	30.22	30.22	30.29	0.00	-0.07
1.60	0.80%	30.58	30.58	30.63	0.00	-0.05
1.75	0.88%	30.94	30.95	30.99	-0.01	-0.05
1.90	0.95%	31.33	31.33	31.37	0.00	-0.04

Table 5.3: Differences in Normal implied volatilities (bps) between our implementation and the results from Antonov's paper, using Input I from Table 5.2

$\frac{K}{F_0}$	K	σ^N	σ^N_A	σ^N_{MC}	diff A	diff MC
-0.95	-0.95%	40.05	40.05	40.86	0.00	-0.81
-0.80	-0.80%	38.43	38.43	39.24	0.00	-0.81
-0.65	-0.65%	36.80	36.80	37.60	0.00	-0.80
-0.50	-0.50%	35.18	35.18	35.97	0.00	-0.79
-0.35	-0.35%	33.59	33.59	34.33	0.00	-0.74
-0.20	-0.20%	32.05	32.05	32.73	0.00	-0.68
-0.05	-0.05%	30.67	30.67	31.25	0.00	-0.58
0.10	0.10%	30.21	30.20	30.63	-0.01	-0.42
0.25	0.25%	30.19	30.19	30.51	0.00	-0.32
0.40	0.40%	30.14	30.14	30.41	0.00	-0.27
0.55	0.55%	30.06	30.06	30.31	0.00	-0.25
0.70	0.70%	30.00	30.00	30.22	0.00	-0.22
0.85	0.85%	29.98	29.98	30.18	0.00	-0.20
1.00	**1.00%**	**30.05**	**30.05**	**30.22**	**0.00**	**-0.17**
1.15	1.15%	30.24	30.24	30.36	0.00	-0.12
1.30	1.30%	30.56	30.56	30.63	0.00	-0.07
1.45	1.45%	31.03	31.03	31.04	0.00	-0.01
1.60	1.60%	31.63	31.63	31.58	0.00	0.05
1.75	1.75%	32.35	32.35	32.26	0.00	0.09
1.90	1.90%	33.17	33.17	33.04	0.00	0.13

Table 5.4: Differences in Normal implied volatilities (bps) between our implementation and the results from Antonov's paper, using Input II from Table 5.2

Parameter	Value
$\mathbf{F}_0$	50 bps
v_0	0.011
γ	0.3
ρ	0
β	0.25
T	3Y

Table 5.5: Input model for figure 5.3

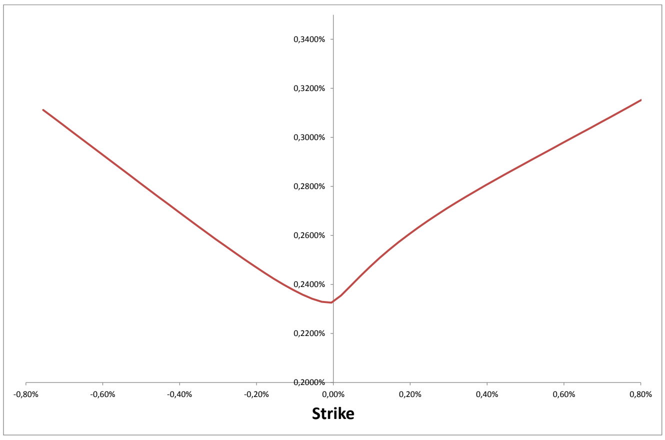


Figure 5.3: Normal implied volatility curve, varying the Strike across zero value

Parameter	Values
$\mathbf{K}$	50 bps
v_0	0.011
γ	0.3
ρ	0
β	0.25
T	3Y

Table 5.6: Input model for figure 5.4

For ϵ small enough.

The figures 5.3 & 5.4 are constructed by respectively using the data used in Tables 5.5 & 5.6. They represent the curve of normal implied volatilities varying respectively through zero-strike and zero-forward. From these we can observe the smoothness of the call price around the zero-strike and the zero-forward. This approach yields coherent results with what we could expect from the expansions 5.30.

5.5 Conclusion

We have seen that this model represents an elegant and natural approach to taking into account the possibility of negative rates. In addition, unlike the Shifted SABR model where we need to postulate a limit on how negative rates can become, this model is free of similar prior adjustments. Thus, it can be directly calibrated to market data.

Unfortunately this model also has its drawbacks. Namely, despite the fact that the zero correlation formula is exact, the mapping technique allowing the parametrization for the general correlation case does not work for all possible parameter combinations. Sometimes this parametrization leads to negative initial volatility. In particular this mapping instability can be very problematic in calibration.

In the next chapter we present an alternative model that was introduced by Antonov to offer an arbitrage free solution.

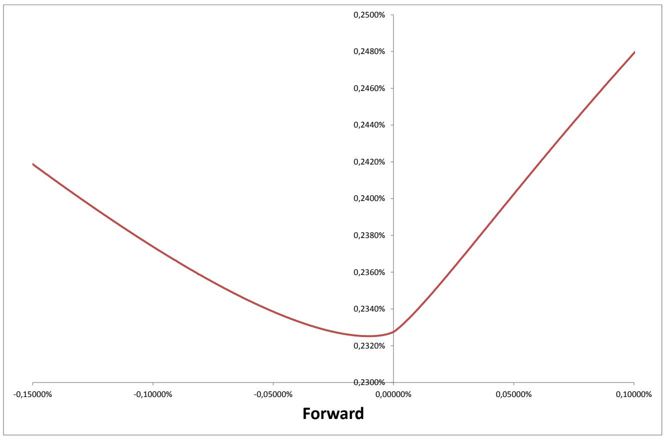


Figure 5.4: Normal implied volatility curve, varying the Forward across zero value

Chapter 6

Mixture SABR

6.1 Introduction

This model was proposed by Antonov in the article Mixing SABR models for Negative Rates [2] as a follow up to the Free Boundary SABR. It came as an alternative solution to deal with the arbitrage problems caused by the approximations in the general correlation Free Boundary SABR.

This new model is a mixture of a zero-correlation Free Boundary SABR and a Normal SABR, and therefore inherits an exact pricing formula for call options, consequently free of arbitrage. Also, as mentioned in the previous chapter, the parametrization for the general correlation can produce incoherent effective parameters. The mixture has no similar problems, and thus can be used in a practical way for calibration and risk management.

In what follows we first present the Normal SABR with a free boundary, before going through the Mixture SABR model.

6.2 Normal SABR

The dynamics of the forward swap rate under the Normal SABR are:

$$dF_t = v_t \, dW_1, \tag{6.1}$$
$$dv_t = \gamma v_t \, dW_2, \tag{6.2}$$

where $dW_1 dW_2 = \rho dt$.

Antonov proposed an exact pricing formula with a 2D integral. The inner integral is the same function G that we have seen in the Free Boundary SABR. Therefore using the approximation for this function reduces the formula to a 1D integral. The time value of a call option is given by:

$$\mathcal{O}_N(T,K) = E[(F_T - K)^+] - (F_0 - K)^+ = \frac{V_0}{\pi}\int_{s_0}^{\infty} \frac{G(\gamma^2 T, s)}{\sinh s}\sqrt{\sinh^2 s - (k - \rho\cosh s)^2} ds \tag{6.3}$$

where

$$\begin{aligned}
\cosh s_0 &= \frac{-\rho k + \sqrt{k^2 + \bar{\rho}^2}}{\bar{\rho}^2}, \\
k &= \frac{K - F_0}{V_0} + \rho, \\
V_0 &= \frac{v_0}{\gamma}, \\
\bar{\rho} &= \sqrt{1 - \rho^2}
\end{aligned}$$

And the function G is given by:

$$G(t,s) = 2\sqrt{2}\frac{e^{-\frac{t}{8}}}{t\sqrt{2\pi t}}\int_s^{\infty} du\, u\, e^{-\frac{u^2}{2t}}\sqrt{\cosh\ u - \cosh\ s} \tag{6.4}$$

We note that this formula is valid for any $\rho \in]-1, 1[$. Also the special cases of zero-strike and/or zero-forward are not problematic here.

Remark. *Careful limits of the integrand should be taken for s around 0. This is the case with At-The-Money options. Also we note that s_0 should always be taken positive.*

6.3 Mixture SABR

Under the SABR Mixture model we assume that the forward rate can be written as:

$$F_t = \chi\, F_t^{(1)} + (1 - \chi)\, F_t^{(2)} \tag{6.5}$$

where

- $F_t^{(1)}$ follows a *zero-correlation* Free SABR model with parameters $(\alpha_1, \beta_1, 0, \gamma_1)$.
- $F_t^{(2)}$ follows a *normal* Free SABR model with parameters $(\alpha_2, 0, \rho_2, \gamma_2)$.
- χ is a random variable taking value 1 with probability p and 0 with probability $1 - p$ and independent of both SABR processes.

Now that we have pricing formulas for both the Free Boundary and the Normal SABR, we can derive the Mixture SABR price.

Assuming the forward rate has dynamics 6.5, we can write:

$$\begin{aligned}
P(F_T \le x) &= P(\chi\, F_T^{(1)} + (1-\chi)\, F_T^{(2)} \le x) \\
&= p\, P(\chi\, F_T^{(1)} + (1-\chi)\, F_T^{(2)} \le x \mid \chi = 1) + (1-p)\, P(\chi\, F_T^{(1)} + (1-\chi)\, F_T^{(2)} \le x \mid \chi = 0) \\
&= p\, P(F_T^{(1)} \le x) + (1-p)\, P(F_T^{(2)} \le x)
\end{aligned}$$

Taking the derivative of the last equation with respect to x yields the probability density function for the mixture model.

$$f_{F_t}(x) = p\, f_{F_t^{(1)}}(x) + (1-p)\, f_{F_t^{(2)}}(x) \tag{6.6}$$

Let's now consider the price of a contingent claim with payoff $h(F_T)$:

$$E[h(F_T)] = \int_{-\infty}^{\infty} h(x) f_{F_T}(x) dx$$

Using 6.6 yields:

$$\begin{aligned} E[h(F_T)] &= p \int_{-\infty}^{\infty} h(x) f_{F_T^{(1)}}(x) dx + (1-p) \int_{-\infty}^{\infty} h(x) f_{F_T^{(2)}}(x) dx \\ &= p\, E[h(F_T^{(1)})] + (1-p)\, E[h(F_T^{(2)})] \end{aligned}$$

Consequently the time value of a call option is given by:

$$\boxed{\mathcal{O}_{Mixture}(T,K) = p\, \mathcal{O}_{ZCFree}(T,K) + (1-p)\, \mathcal{O}_{Normal}(T,K) \qquad (6.7)}$$

Since we have free of arbitrage exact formulas for $\mathcal{O}_{ZCFree}(T,K)$ and $\mathcal{O}_{Normal}(T,K)$, the call price under the Mixture SABR model is consequently free of arbitrage.

Unlike the original SABR model where there are 4 degrees of freedom, the Mixture model has 7. It was pointed out in [2] that these new degrees offer a new possibility of a joint calibration of swaptions and CMS payments. Antonov proposed two ways (full and reduced) of choosing parameters for calibration purposes. In both, the initial stochastic volatility was set for the Free and Normal SABR models at the ATM volatility σ_0:

$$\sigma_0 = \alpha_1 F_0^{\beta_1} = \alpha_2 \tag{6.8}$$

In the full parametrization the probability p is taken as a function of a parameter x, that should then be calibrated:

$$p(x) = \frac{\sigma_0\, \beta_1\, e^x}{\sigma_0\, \beta_1\, e^x + |\gamma_2\, \rho_2|} \tag{6.9}$$

The reduced parametrization corresponds for the probability to $x = 0$

$$p = \frac{\sigma_0\, \beta_1}{\sigma_0\, \beta_1\ + |\gamma_2\, \rho_2|} \tag{6.10}$$

and is fully determined by the other model parameters $(\beta_1, \gamma_2, \rho_2)$.

Furthermore, the vol-of-vol are linked between the two models in the reduced parametrization:

$$\gamma_2 = \frac{\gamma_1}{1-\beta_1} \tag{6.11}$$

As mentioned above, Antonov explained that the extra degrees of freedom in the full parametrization, offers the possibility of a joint calibration to swaptions and CMS payments, we however focus on calibration to swaptions alone.

Furthermore the model will be used in calibrating the swaption cube, which in our study consists of approximately 500 "smiles". As a result, we need the calibration to be efficient and relatively fast. We recall that the pricing formula includes 2D integrals or 1D integrals in the approximated version. Thus, the pricing process is much heavier than Hagan's approximation. And having more degrees of freedom, means more calibration parameters and consequently more execution time.

We will therefore not be using the full parametrization. We further discuss our choices in section 6.4.2.

6.4 Numerical Experiments

In this section we present the numerical results obtained using our implementation of the Mixture SABR model. The experiments include comparison with some results from the original article in addition to a large spectrum of experiments that we have done.

6.4.1 Mixture SABR PDF

The implied probability density function is very important in this study. The Mixture SABR model came as a new way of going around the arbitrage problems of the mapping technique, therefore we expect it to inherit the behavior of the Free Boundary pdf, in particular we would like it to reproduce the stickiness of the forward rate around zero. In this section we use the results derived in section 5.4.1.

Let us consider the input from Table 6.1.

Parameter	Input I	Input II
F_0	50 bps	50 bps
α_1	0.6 $F_0^{1-\beta_1}$	0.6 $F_0^{1-\beta_1}$
β_1	**0.1**	**0.25**
γ_1	0.3	0.3
α_2	$\alpha_1 F_0^{\beta_1}$	$\alpha_1 F_0^{\beta_1}$
γ_2	$\frac{\gamma_1}{1-\beta_1}$	$\frac{\gamma_1}{1-\beta_1}$
ρ_2	-0.3	-0.3
p	0.5	0.5
T	3Y	3Y

Table 6.1: Implied probability density function inputs

The Figures 6.1 & 6.2 show that the Mixture SABR does inherit the desired behavior. Also the parameter β_1 still controls the intensity around zero. We can deduce that this model reproduces forward dynamics similar to the Free Boundary's, with the advantage of being arbitrage free.

Parameter p effect:

Let us now take a closer look at the probability p effect. This parameter allows to navigate between the zero-correlation Free boundary ($p = 1$) and the Normal SABR ($p = 0$). Figures 6.3 & 6.4 represent the PDF of the inputs in table 6.1 for different values of p.

We observe that the parameter p controls the intensity around 0, this was predictable since by lowering p we lower the contribution of the ZC Free Boundary, and consequently the effect of β_1. The effect of the correlation is less visual on these but undoubtedly present, in fact it can be observed by plotting the PDF of the Normal SABR with $\rho = 0$ and $\rho \neq 0$ (Figure 6.5). Thus the correlation seems to affect the tails.

Mixture PDF restrictions

Even though this model seems capable of reproducing the Free Boundary's density, there are some restrictions. Indeed the contribution of the correlation in the mixture model is only coming from the normal SABR, thus it would be a problem if for some reason we need to consider extreme values for this parameter (e.g. $\rho = 1$).

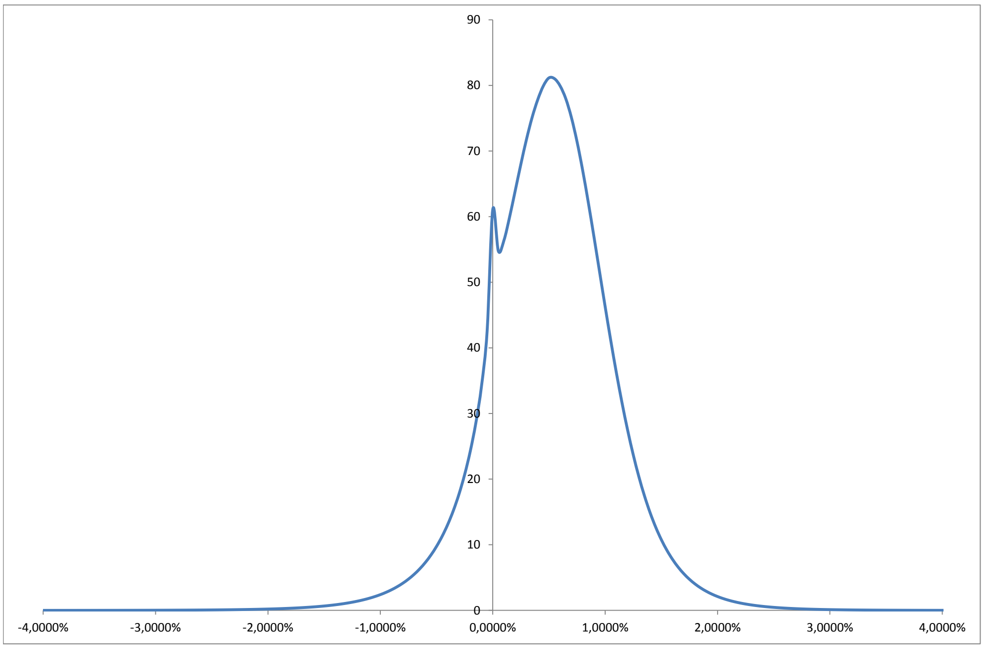


Figure 6.1: Mixture SABR implied PDF for Input I: $\beta = 0.1$

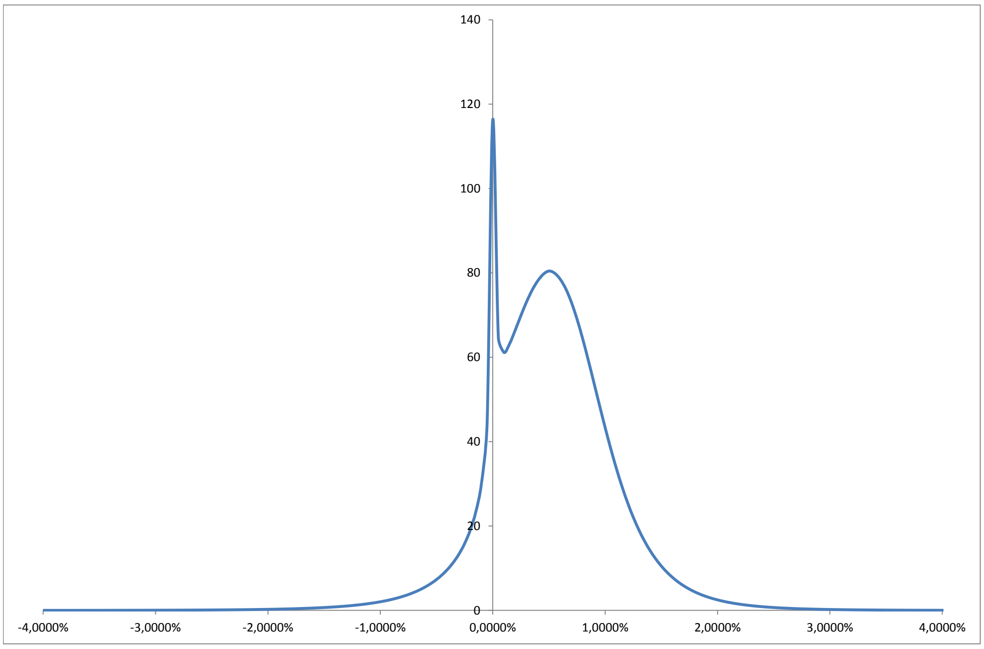


Figure 6.2: Mixture SABR implied PDF for Input II: $\beta = 0.25$

Unless $p = 0$, its contribution can only be partial. Therefore, we expect that there would be some distribution shapes of the general Free boundary that cannot be reproduced. The same reasoning applies to the parameter β, it is however less problematic since we will not need to reach high values (i.e. near 0.50) in practice, and $\beta = 0$ corresponds to the normal model.

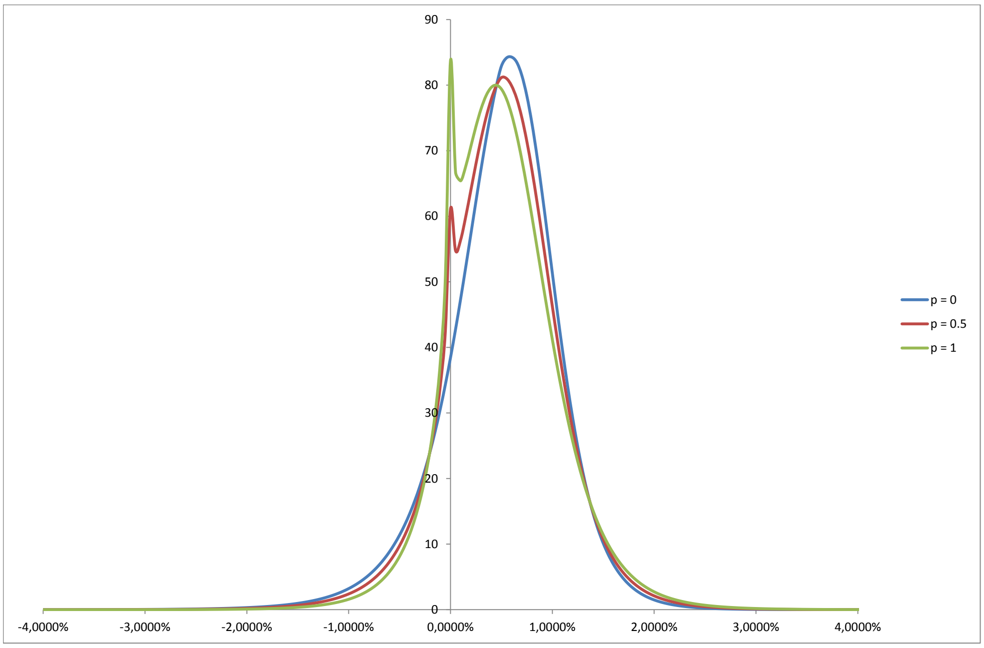


Figure 6.3: Mixture SABR implied PDF for Input I: $\beta = 0.1$. With $p \in \{0, 0.5, 1\}$

Left tail

Shifting the SABR model with a fixed $s > 0$, implies that $P^{\text{Shifted}}(F < -s) = 0$. Since the shift is fixed prior to calibration its value doesn't reflect any specific market behavior, and comes only from our decision of how negative rates can become. This is one of the main reasons of considering an alternative approach.

$P^{\text{Mixture}}(F < -s)$ for the Mixture SABR is strictly positive , however we want to quantify it in order to get a hold of how permissive this model can be. In addition it is important to study this behavior when varying the parameter p, as this allows the comparison of the left tail between the Mixture (with $p \neq 0$) and the full Normal SABR. As seen in Figure 6.5 the correlation of the Normal SABR has a significant effect on the left tail, hence the importance to perform the comparison on 3 different cases: $\rho \in \{-0.98, 0, 0.98\}$.

Figure 6.6 shows the distribution function valued at -2% ($P(F < -2\%)$) as a function of the probability p. We see that the Normal SABR has a fatter tail than the zero-correlation Free Boundary, however the slope is reverted when considering high enough correlation values. We note that when $F_0 < 0$ the slope is reverted for negative correlation values.

These results are in line with the correlation effect observed in Figure 6.5 and shows that the behavior of the left

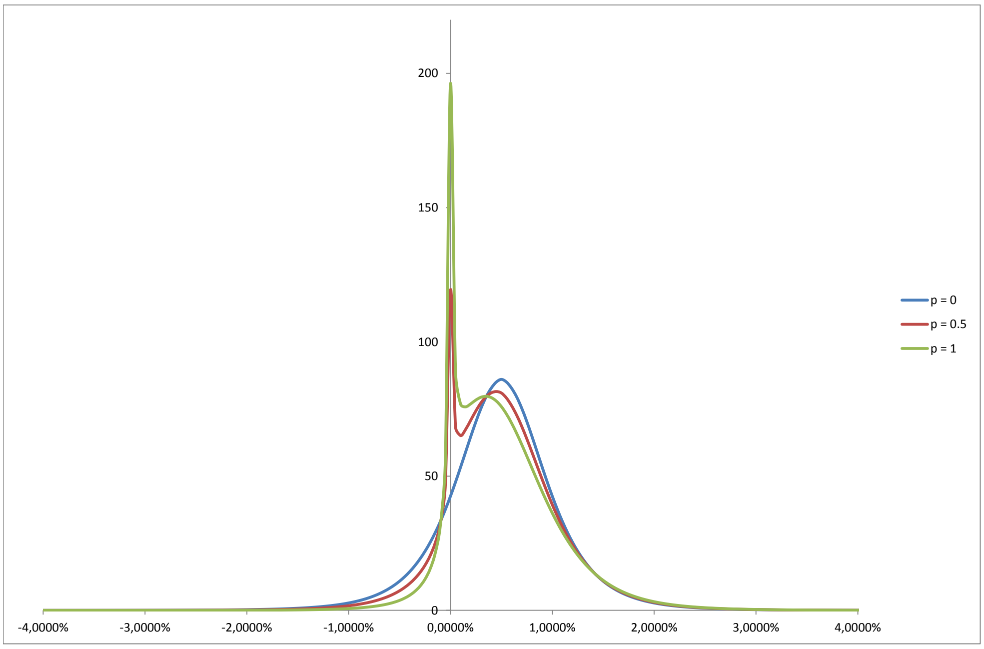


Figure 6.4: Mixture SABR implied PDF for Input II: $\beta = 0.25$. With $p \in \{0, 0.5, 1\}$

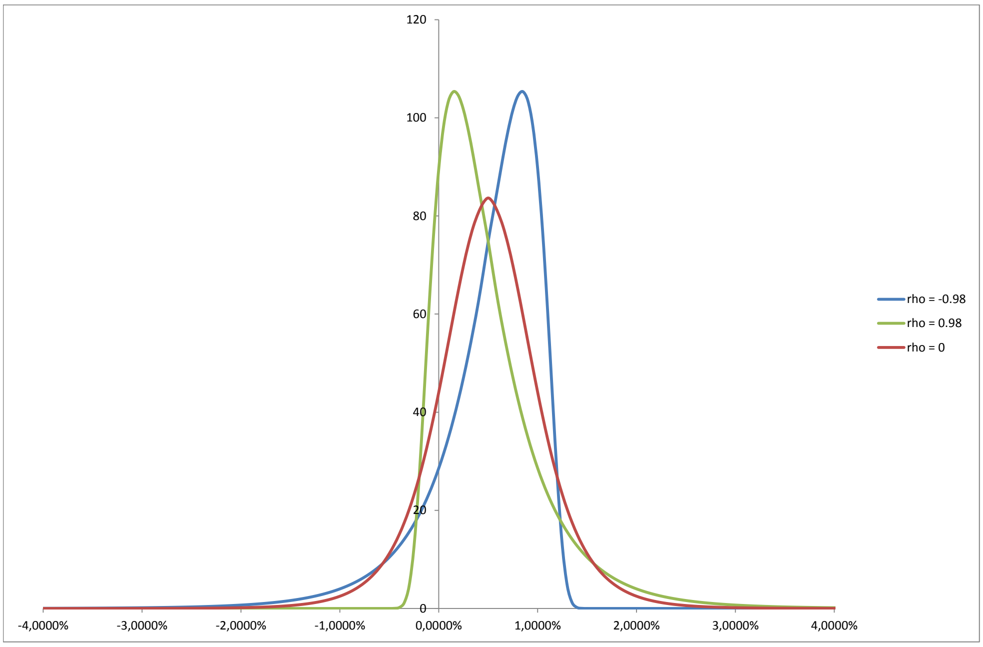


Figure 6.5: Normal SABR PDF using table 6.1, with probability $p = 0$. $\rho_2 = 0$ (red), $\rho_2 = -0.98$ (blue) and $\rho_2 = 0.98$ (green)

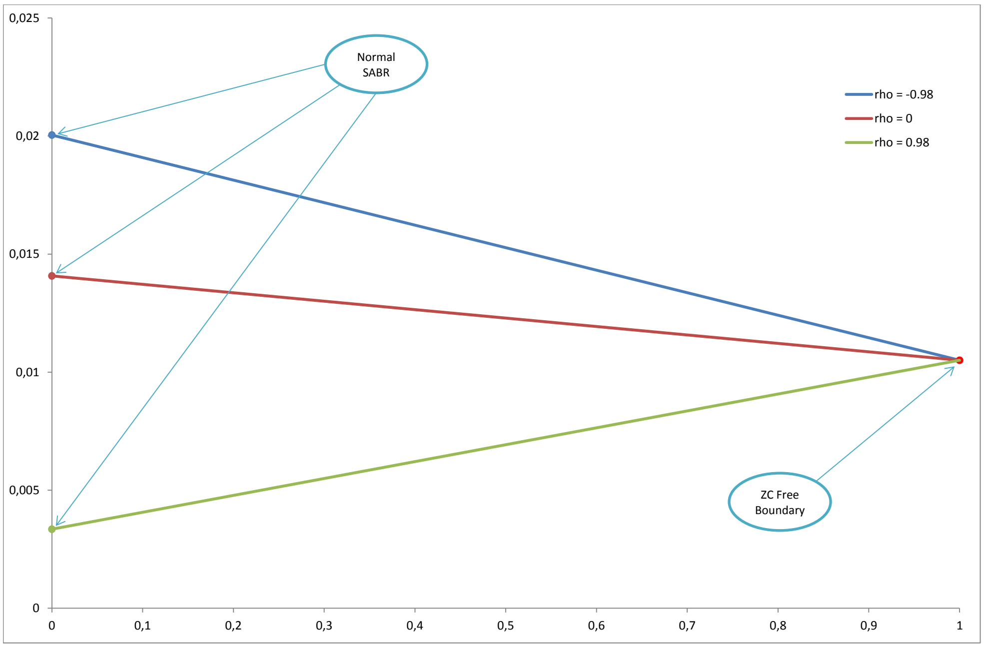


Figure 6.6: Plot of $P(F < -2\%)$ as a function of the probability p of the Mixture SABR Model

tail can be problematic when calibrating to market data. Indeed we observe that the Shifted SABR fits almost perfectly the market and since its implied distribution is equal to zero for $F < -s$, we expect the mixture to have some difficulties in producing the same fit quality on both positive and negative strikes. This point is further discussed in the next section.

6.4.2 Calibration

In this part we go through the different steps of calibrating the Mixture SABR model, we start by presenting the general approach and work our way through the complications we have met and the choices we have made. This is an important part of this study, as it gives the practical usefulness of the model, and allows us to assess its performance.

General approach

Calibration of the swaption cube, is performed by calibrating the model to every single smile, i.e. a set of volatilities/prices corresponding to different strikes for a fixed maturity and tenor . As a result, for each couple $(\text{Maturity}_i, \text{Tenor}_j)$ there will be a corresponding fitted Mixture SABR model. Thus, we only focus on calibrating the model to a single smile, the same procedure is then reiterated to the rest of the cube.

Let's start by fixing some notations.

- T : Maturity of the swaption
- T_n: Tenor (see page 6)
- n : number of the available swaption quotes
- K_i: Strikes ($i \in \{1, .., n\}$)
- σ_i^N : Normal implied volatility

- $\sigma_i^{shifted}$: shifted BS implied volatility
- s : shift corresponding to $(\sigma_i^{shifted})_i$.

The Mixture SABR is calibrated using prices, thus if we have shifted or normal implied volatilities as input, we transform them to prices. Either way, in order to assess the quality of the fit we use implied volatilities.

The main goal is to calibrate the model to swaptions. As this is a vanilla product, we need the calibration to be relatively fast. Therefore we can not afford to calibrate using all the parameters, instead we use a reduced version. In fact, we use a slightly different parametrization from what was proposed in [2]. We set the initial volatilities such that:

$$\alpha_2 = \alpha_1 |F_0|^{\beta_1}$$

However we do not choose them to be equal to the ATM normal implied volatility. Instead they will be fixed by fitting the ATM price, this is detailed in the calibration procedure.

In the reduced parametrization suggested by Antonov:

$$p = \frac{\beta_1 \sigma_0}{\beta_1 \sigma_0 + |\rho_2 \gamma_2|}$$

σ_0 is the ATM normal implied volatility.

This creates a difference in scale between the two terms $\beta_1 \sigma_0$ and $|\rho_2 \gamma_2|$ as σ_0 is too small (around 1%) in comparison to β_1, ρ_2 and $gamma_2$. Thus unless ρ_2 is equal to zero the first term will always be negligible. Consequently p will be equal to 1 in the zero correlation case, and otherwise approximately equal to 0. This means that in most cases, it will be as if we only use the Normal SABR model.

The parametrization below circumvents this problem while keeping a coherent set up. When β_1 is equal to zero p is also equal to zero, and when ρ_2 is equal to zero, p is equal to 1.

$$p = \frac{\beta_1}{\beta_1 + |\rho_2|} \tag{6.12}$$

Finally we link the vol-of-vols as suggested in Antonov's reduced parametrization.

$$\gamma_2 = \frac{\gamma_1}{1 - \beta_1} \tag{6.13}$$

Consequently we reduce the calibration parameters as follows:

$$(\alpha_1, \beta_1, \gamma_1, \alpha_2, \rho_2, \gamma_2) \to (\alpha_1, \beta_1, \gamma_1, \rho_2)$$

We have the same calibration set up as the classical SABR model. We perform the calibration following these steps:

- Fix β_1 in {0.1,0.2,0.3}
- Calibrate using the parameters (ρ_2, γ_1) on a suitable range as described below. For each couple of vol-of-vol and correlation α_1 is set by fitting the ATM price
- Choose the fit corresponding to the β_1 yielding the smallest fit error of a given norm.

By proceeding as described above, we ensure a perfect fit ATM. This is important for us, since these options are the most liquid and we want the model to perfectly reproduces there prices.

At-The-Money fit

This part details the ATM fit. We note that at this point β_1^0, γ_1^0 and ρ_2^0 are assumed to be fixed.

The problem can be formulated in two ways, as a minimization of an objective function, or as a root finding problem.

$$\alpha_1^0 = \text{Arg}\left(\min_{\alpha_1 \in \mathcal{A}} ||V^{\text{Mixture}}(\alpha_1, \beta_1^0, \gamma_1^0, \rho_2^0) - V_{\text{ATM}}^{\text{Market}}||^2 \right) \tag{6.14}$$

where $||.||$ is a norm, and $\mathcal{A}$ the set of possible values of α_1.

or,

$$\alpha_1^0 = \text{Arg}\left(V^{\text{Mixture}}(\alpha_1, \beta_1^0, \gamma_1^0, \rho_2^0) - V_{\text{ATM}}^{\text{Market}} = 0 \right) \tag{6.15}$$

We prefer the latter as we require a perfect ATM fit. This however introduces a theoretical obstacle, indeed, depending on the values of γ_1^0 and ρ_2^0 equation 6.15 may not have a solution. Nevertheless, we can also see this as an opportunity to reduce the range of calibration.

We note that the same difficulties can be met when calibrating Hagan's formula. Indeed the ATM fit is done by solving in α the following 3rd degree equation:

$$\sigma_{ATM} = \sigma_B(f, K) = \frac{\alpha}{f^{(1-\beta)}} \left\{ 1 + \left[\frac{(1-\beta)^2}{24} \frac{\alpha^2}{f^{(2-2\beta)}} + \frac{1}{4} \frac{\rho\beta\alpha\gamma}{f^{(1-\beta)}} + \frac{2-3\rho^2}{24}\gamma^2 \right] T \right\}$$

This equation has a at least one real solution, however this solution is not necessarily positive. Thus it is possible to have a set of parameters of the correlation and the vol-of-vol that are not valid. In fact, if we set a calibration space $\mathcal{C} = [\rho_{\min}, \rho_{\max}] \times [\gamma_{\min}, \gamma_{\max}]$, we find out that there is a whole region where we don't find a positive solution, however formally characterizing it is somewhat difficult.

For the Mixture SABR we do not have an analytical equation, which makes it even more difficult to understand if such behavior occurs. In order to get an intuition, we proceed with a numerical experiment.

To define a clear framework, we use a specific example. However the stated results are general as shown following the example. Let us consider the market data on the 24/10/2016 in Table 6.2.

We discretize the space $\mathcal{C}$ into $\mathcal{C}_d = \{(\rho_i, \gamma_j), (i, j) \in [\![1, n]\!] \times [\![1, m]\!]\}$, and for now we only consider

Parameter	Value
F_0	1.383%
σ_{ATM}^N	0.6635%
T	10Y
T_n	10Y

Table 6.2: ATM Swaption 10Y10: normal implied volatility and forward observed on the 24/10/2016

the Normal SABR ($p = 0$). For each point in $\mathcal{C}_d$ we try to fit the ATM on $\alpha \in [\alpha_{\min}, \alpha_{\max}]$. We note that $\alpha_{\min}$ should be taken strictly positive, otherwise the volatility would always be equal to zero, thus leading to a zero forward rate (see the dynamics 5.1).

The Figure 6.7 shows the regions where the fit succeeded (green) and where it did not (red). The process of fitting the ATM formalized in equation 6.15 starts by bracketing the root, i.e. finding an α_{lw} and α_{up} such

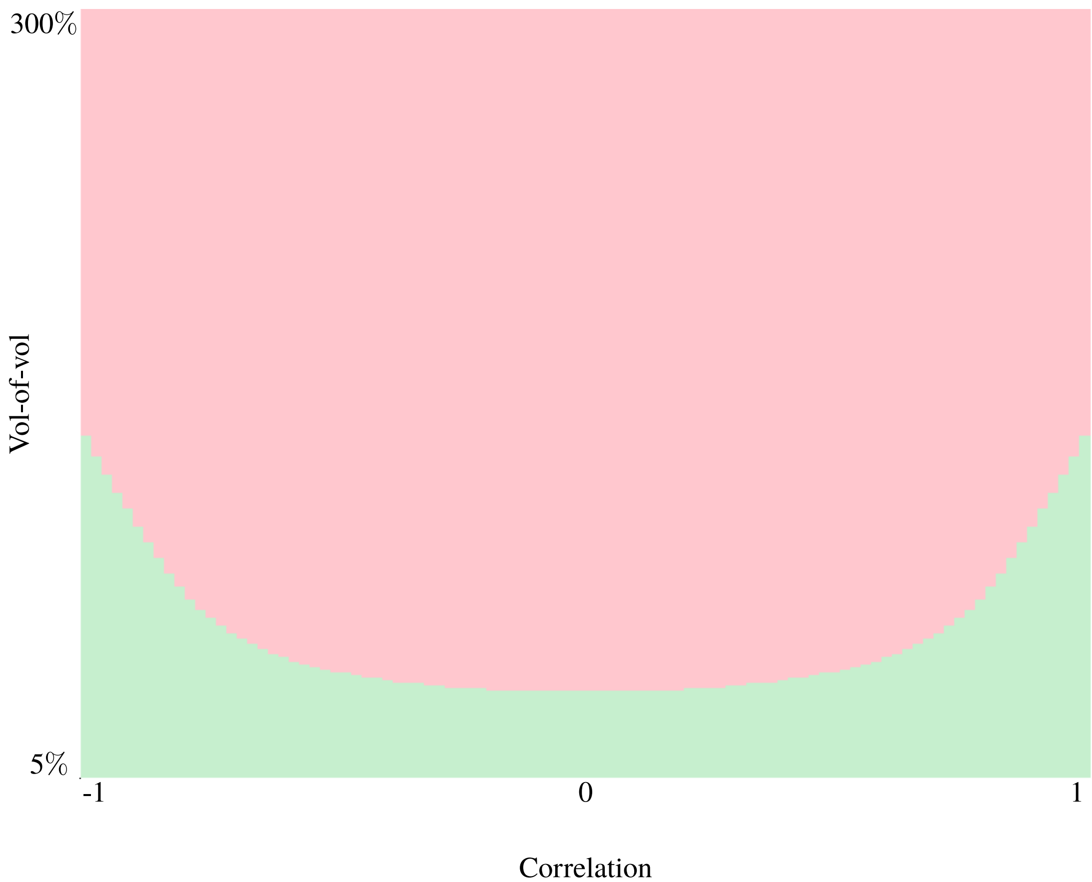


Figure 6.7: ATM fit of the Normal SABR for each point in $\mathcal{C}_d$, given $\alpha_{\min} = 1\%$. The green region represents the points where the fit succeeded, whereas the red region is where it failed

that:

$$\left(V^{\text{Mixture}}(\alpha_{\text{lw}}, \beta_1^0, \gamma_1^0, \rho_2^0) - V_{\text{ATM}}^{\text{Market}}\right)\left(V^{\text{Mixture}}(\alpha_{\text{up}}, \beta_1^0, \gamma_1^0, \rho_2^0) - V_{\text{ATM}}^{\text{Market}}\right) < 0$$

Where $\alpha_{\text{lw}}, \alpha_{\text{up}} \in [\alpha_{\min}, \alpha_{\max}]$, and $\beta_1^0 = 0$.

Having a closer look at the red region, we find out that when we cannot fit the ATM value, we always have:

$$V^{\text{Mixture}}(\alpha, \beta_1^0, \gamma_1^0, \rho_2^0) > V_{\text{ATM}}^{\text{Market}}, \text{for every } \alpha \in [\alpha_{\min}, \alpha_{\max}]$$

Therefore there is a set of correlation and vol-of-vol values for which the price given by the model is always greater than the market price for any value of α. This means that if we want to calibrate the smile in this example it is pointless to consider the same range. Thus it is important to characterize the border line, and consequently reduce the calibration space.

If we fix ρ and look at the range of permissive values of γ, we can see from Figure 6.7 that worst case corresponds to $\rho = 0$, this suggests that we obtain the highest prices in the zero-correlation case. And taking into account the symmetric shape of the boundary line with respect to ρ, we postulate the following proposition:

Proposition:
The ATM time value under the Normal SABR as a function of the correlation has a maximum at $\rho = 0$.

Proof. The ATM time value under the Normal SABR is given by:

$$\begin{aligned}\mathcal{O}_N(T, F_0) &= E[(F_T - F_0)^+] \\ &= \frac{v_0}{\pi\gamma}\int_0^\infty \frac{G(\gamma^2 T, s)}{\sinh s}\sqrt{\sinh^2 s - \rho^2(1-\cosh s)^2}\, ds\end{aligned}$$

Taking the derivative with respect to the correlation yields:

$$\frac{\partial \mathcal{O}_N}{\partial \rho}(T, F_0) = (-\rho)\,\frac{v_0}{\pi\gamma}\int_0^\infty \frac{G(\gamma^2 T, s)}{\sinh s}\frac{(1-\cosh s)^2}{\sqrt{\sinh^2 s - \rho^2(1-\cosh s)^2}}\, ds$$

Thus, the time value as a function of ρ is increasing on $[-1, 0]$ and decreasing on $[0, 1]$. Therefore it reaches its maximum at $\rho = 0$.

$\square$

This result coupled with the fact that the prices obtained in the red region (see Figure 6.7) are higher than the market price means that if we decide to calibrate on a rectangle $\subset \mathcal{C}$ by reducing the value of $\gamma_{\max}$, it is then possible to obtain a value $\gamma_{\text{opt}}^{\max}$ for which we can fit the ATM price for any $\gamma < \gamma_{\text{opt}}$. This value corresponds to the boundary point between the red and green regions for $\rho = 0$.

Now if we can determine upfront to the calibration process the value of γ_{opt}, we would be certain to consider a valid calibration space where it is always possible to obtain a "perfect" ATM fit. This part can in fact be studied for the Mixture SABR with a fixed $p \neq 0$, because adding the zero-correlation Free Boundary only shifts the border line up or down without changing its shape, this is logical since by construction the Free Boundary is independent of the parameter ρ, as illustrated on Figure 6.8.

However if we consider a parametrization for the probability p that depends on the correlation, we can obtain behaviors that can be difficult to characterize. This is the case for the parametrization given in equation 6.12. In order to avoid such complications, we instead find γ_{opt} for both the zero-correlation Free Boundary and the Normal SABR and then choose $\min(\gamma_{\text{opt, N}}, \gamma_{\text{opt, F}})$ where $\gamma_{\text{opt, N}}$ $(\gamma_{\text{opt, F}})$corresponds to γ_{opt} with $p = 0$

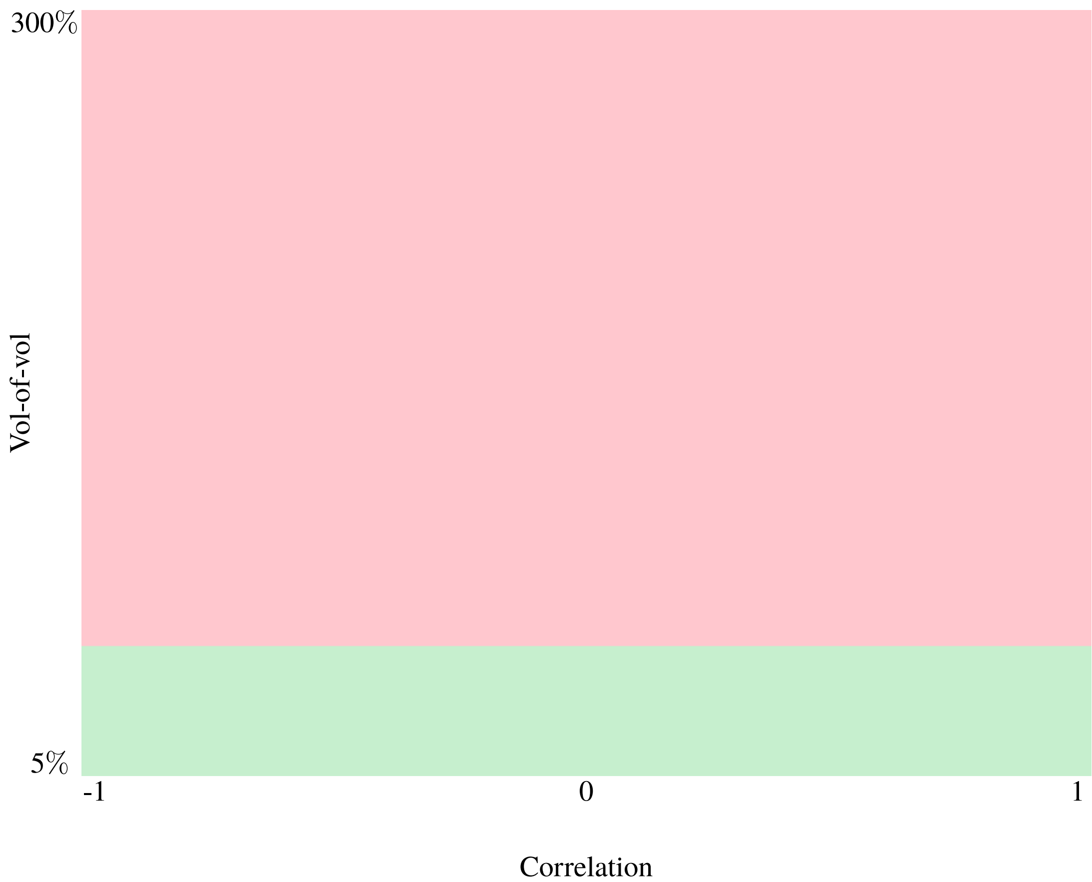


Figure 6.8: ATM fit of the zero-correlation Free Boundary for each point in $\mathcal{C}_d$, given $\alpha_{\min} = 1\%$. The green region represents the points where the fit succeeded, whereas the red region is where it failed

$(p = 1)$, . Since the Mixture price is a weighted average of the ZC Free Boundary and the Normal SABR, then if we set the value of vol-of-vol like that, we will be sure that both models are capable of reaching the market price, and so will their weighted average no matter how the weights are distributed.

For both models, we observe that the value of γ_{opt} is determined by the lower boundary that we set for $\alpha_{\min}$, in fact the price is an increasing function of the initial volatility and the vol-of-vol. Thus an increase of the value of the vol-of-vol needs to be compensated by a decrease of the initial vol. And since it doesn't make any sense to consider a zero initial vol, we need to set a reasonable lower boundary, consequently implying an upper boundary for the vol-of-vol.

Therefore the optimal value of the vol-of-vol is obtained as the maximum value for which it is possible to fit the ATM price, given that $\alpha_{\min}$. This can be formalized by:

$$\gamma_{\text{opt, N}} = \text{Arg}\bigg(V^{\text{N}}(\alpha_{\min}, 0, \gamma, 0) - V^{\text{Market}}_{\text{ATM}} = 0\bigg) \tag{6.16}$$

$$\gamma_{\text{opt, F}} = \text{Arg}\bigg(V^{\text{F}}(\alpha_{\min}, \beta^0, \gamma, 0) - V^{\text{Market}}_{\text{ATM}} = 0\bigg) \tag{6.17}$$

Applying this process prior to the calibration guarantees a calibration space where it is always possible to fit the ATM price. Therefore the pre-calibration goes as follow:

- Choose the initial boundaries: $\alpha_{\min}, \alpha_{\max}, \gamma_{\min}, \gamma_{\max}$.
- Find $\gamma_{\text{opt, N}}$ and $\gamma_{\text{opt, F}}$ by numerically solving the equations 6.16 & 6.17.
- Set $\gamma_{\text{opt, M}} = \min(\gamma_{\text{opt, N}}(1 - \beta_1); \gamma_{\text{opt, F}})$
- The resulting calibration space for the Mixture SABR is then:
  $\mathcal{C} = \{\ (\rho, \gamma)\ /\ \rho \in [-0.98, 0.98] \text{ and } \gamma \in [\gamma_{\min}, \gamma_{\text{opt, M}}]\ \}$

This set is then used as a calibration input as described in the next section.

Smile calibration

In order to calibrate the Mixture SABR model to a swaption smile, we have used the previously specified parametrization. Thus allowing us to reduce the parameters to $(\alpha_1, \beta_1, \gamma_1, \rho_2)$. The β_1 is fixed, α_1 is obtained by fitting the ATM price. Therefore the calibration parameters are (γ_1, ρ_2). A suitable range for γ_1 is chosen following the steps in the previous section.

This two dimensional calibration is performed using the **Downhill Simplex Method** by minimizing an objective function over the space $\mathcal{C}$:

$$\min_{(\rho_2, \gamma_1) \in \mathcal{C}} \sum_{i=1}^{n} ||V^{\text{Mixture}}(\alpha_1, \beta_1, \gamma_1, \rho_2, K_i) - V^{\text{Market}}(K_i)||^2$$

Where $||.||$ is a chosen norm.

In this section we assess the quality of calibration to market data of this model, discuss its speed and analyze the fitted parameters.

We use market shifted Black implied volatilities (i.e the price is obtained using a shifted log normal model) for the swaption 10Y10 observed at the 24/06/2016 for the EUR currency. The shift used here is $s = 3\%$, and the spot swap rate is $F_0 = 1.35\%$.

We calibrate the Mixture and Shifted SABR model to this data. The Mixture is calibrated using the reduced

Relative Strike(%)	Strike(%)	Vol(%)
-2.50	-1.15	20.7
-2.00	-0.65	19.7
-1.50	-0.15	19.7
-1.00	0.35	17.9
-0.50	0.85	17.2
0	1.35	16.7
0.50	1.85	16.3
1.00	2.35	15.9
1.50	2.85	15.5
2.00	3.35	15.2
2.50	3.85	14.9

Table 6.3: Swaption 10Y10 calibration input. The strike is equal to the sum of the relative strike and the spot swap rate.

parametrization specified in this chapter, in addition to a perfect ATM fit. The Shifted is calibrated using a similar approach. The calibrated parameters are presented in Table 6.4.

Table 6.5 and Figure 6.9 present the shifted implied volatilities obtained from the calibrated models. Overall

Shifted		Mixture	
α	0.1725	α_1	0.0111
β	1	β_1	0.1
γ	0.1453	γ_1	0.0786
ρ	-0.6269	α_2	0.0072
s	3.00%	γ_2	0.0873
		ρ_2	0.98
		p	0.0926

Table 6.4: Mixture and Shifted SABR calibrated to the data in Table 6.3

Strike(%)	Market(%)	Shifted(%)	Mixture(%)
-1.15	20.7	20.8	21.4
-0.65	19.7	19.6	19.8
-0.15	19.7	18.7	18.7
0.35	17.9	17.9	17.9
0.85	17.2	17.3	17.2
1.35	16.7	16.7	16.7
1.85	16.3	16.2	16.3
2.35	15.9	15.8	15.9
2.85	15.5	15.5	15.5
3.35	15.2	15.2	15.2
3.85	14.9	14.9	15.0

Table 6.5: Calibrated shifted volatilities

the fit is good for both models, however the Mixture SABR tend to yield higher volatilities on extreme negative strikes.This behavior is not an isolated case and is observed on many examples.

The Shifted SABR fits almost perfectly the smile, thus it is reasonable to suggest that its dynamics are aligned with the market. And since the call price is an integral over $[-\infty, K]$ and the implied density of the Shifted is

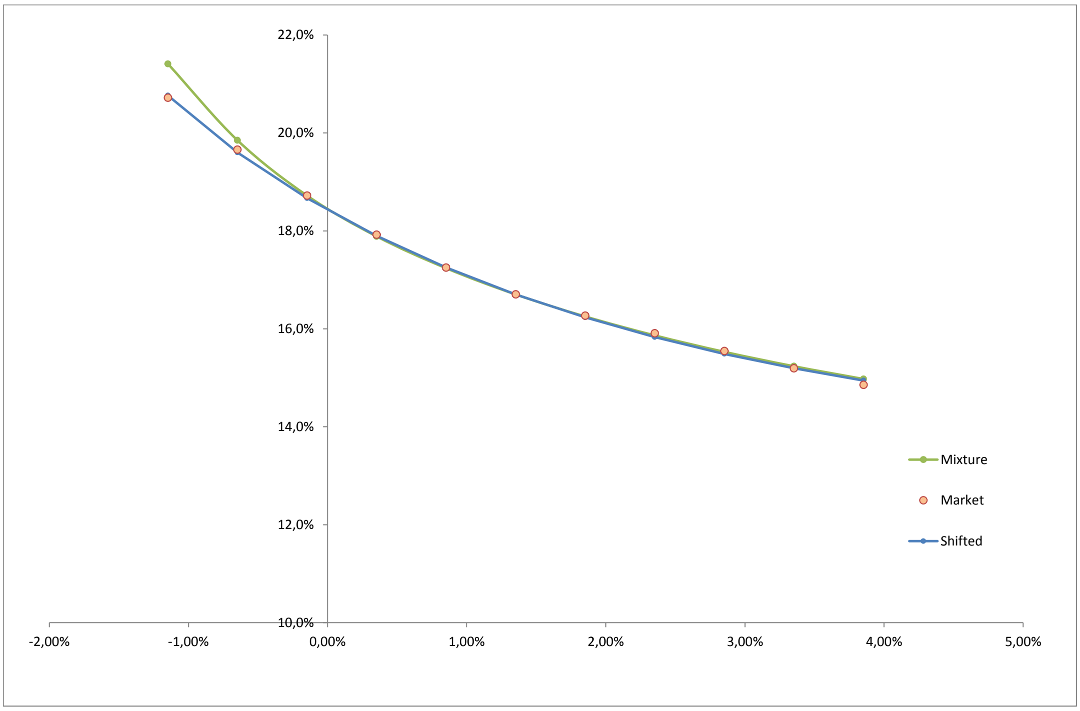


Figure 6.9: Calibrated Shifted and Mixture SABR shifted implied volatility corresponding to the input in Table 6.3

equal to zero over $[-\infty, -s]$, the prices yielded by the calibrated shifted on very far OTM options with negative strikes are low compared with those yielded by the fitted Mixture SABR.

If we take a closer look at the Mixture fitted parameters, we can see that the correlation is at its maximum value "0.98", and based on the discussion around the Mixture PDF restrictions in 6.4.1, it can be explained by the fact that this parameter controls the thickness of the tails. At high positive correlation values the left tail tends to be flattened, consequently reducing the prices on negatives strikes. Therefore we can conclude that this reduced parametrization of the Mixture cannot perfectly fit some smiles, as we expect the calibrated model to take off from the shifted on the left region.

Having said that, this extreme OTM points are not very liquid and might simply be obtained by extrapolating the Shifted SABR, thus explaining its perfect fit.

Based on this, if we decide to include such extreme points in calibration, we will usually end up with a correlation around 0.98 when $F_0 \geq 0$ (or -0.98 when $F_0 < 0$). This means that calibration would be reduced to one parameter. Furthermore as ρ takes extreme values the probability p becomes low, consequently leading to a small contribution of the zero-correlation Free Boundary.

In order to be comfortable around the calibration results, we chose to plot the RMSE(Root-mean-square error) surface: we discretise $\mathcal{C}$ and compute the RMSE on each point. This way it is possible to verify whether the "Simplex" calibrated model corresponds to a global minimum or just a local minimum.

Figure 6.10 shows that the calibrated model corresponds indeed to a global minimum. In fact while taking into account the extreme left points in calibration, we usually obtain similar surfaces, particularly the minimum is obtained for high values of correlation.

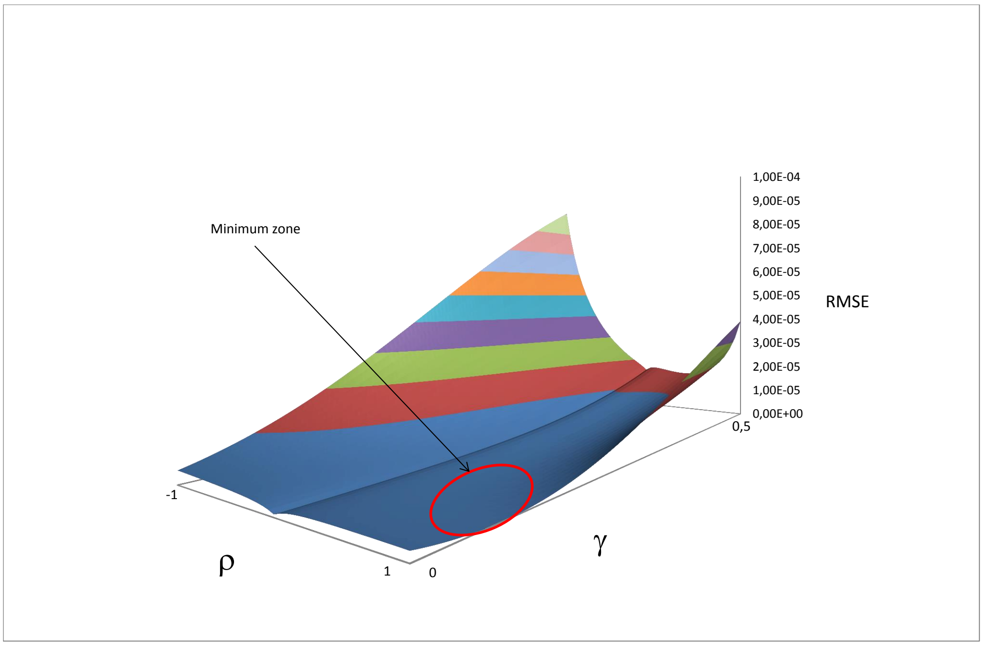


Figure 6.10: RMSE(Root-mean-square error) surface for the Mixture SABR on a subset of $\mathcal{C}$ for input in Table 6.3

Another way to look at this is to see how the smile moves when we reduce the correlation. Figure 6.11 shows that as we reduce it the left values go further up. Figures (6.12, 6.13 and 6.14)show the smile alteration when changing the different reduced model parameters.

At this point, the decision of using this model depends on our confidence in the quality of these extreme left points. We can distinguish two setups:

- We suppose that these points should not be taking into account in the calibration. The model can then be used with this parametrization and is likely to produce higher prices on that region.
- We suppose that these points are valid. The model with this parametrization is then not suitable, in this case we should either use the Shifted or increase the Mixture calibration degrees of freedom. This last option is likely to offer more flexibility and possibly allow to fit the whole smile, however the calibration process is going to be very slow.

Swaption cube calibration

Once it is possible to calibrate the model to a smile, the swaption cube calibration becomes very easy. Indeed, for each couple (T, T_n) we calibrate the corresponding smile.

The difficulty at this point is execution time, we want the calibration to be as fast as it can be. In our study we have 513 couples (T, T_n), thus if calibrating a smile takes 1 second calibrating the cube will take 8.55 minutes.

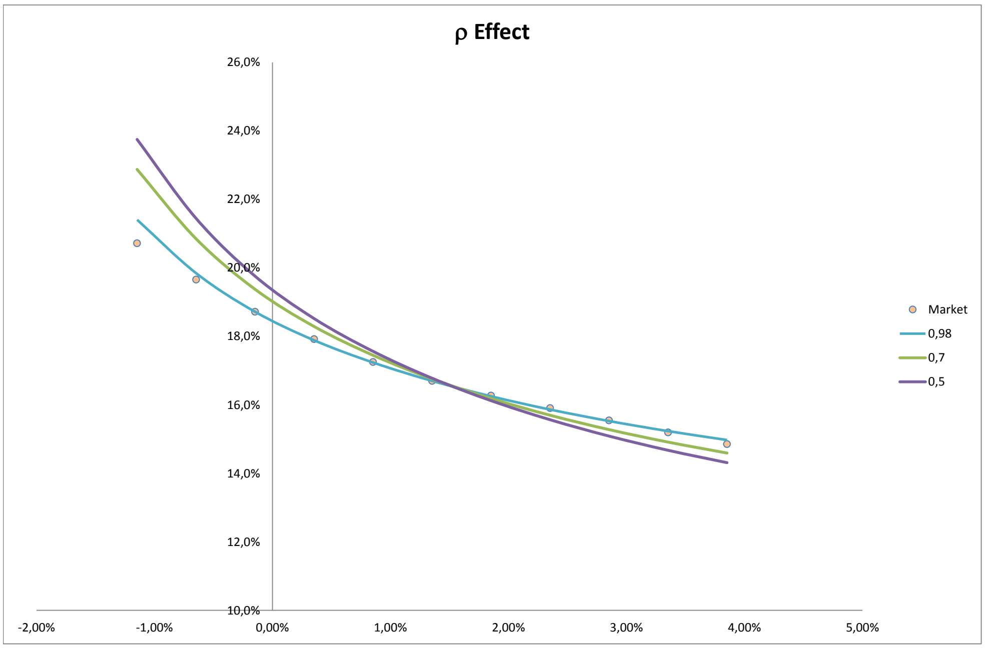


Figure 6.11: Correlation effect for input in Table 6.3

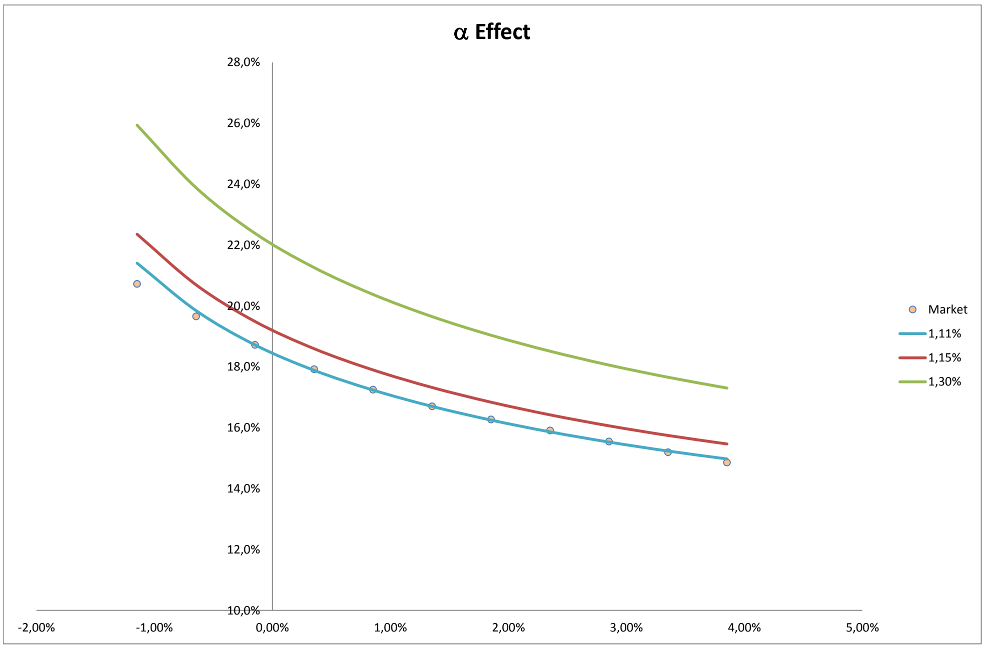


Figure 6.12: Alpha effect for input in Table 6.3

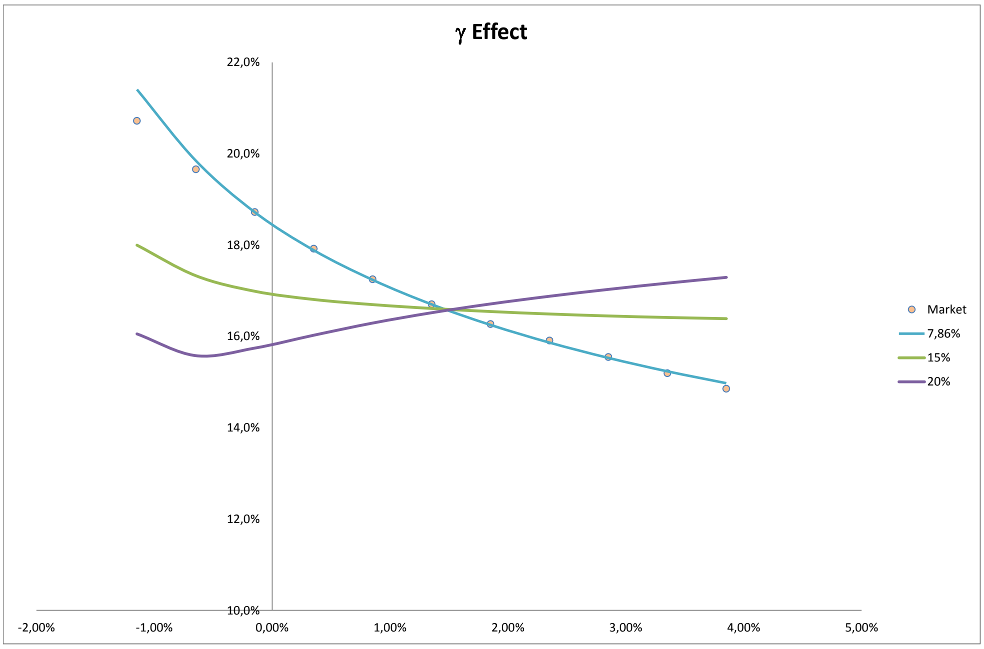


Figure 6.13: Gamma effect for input in Table 6.3

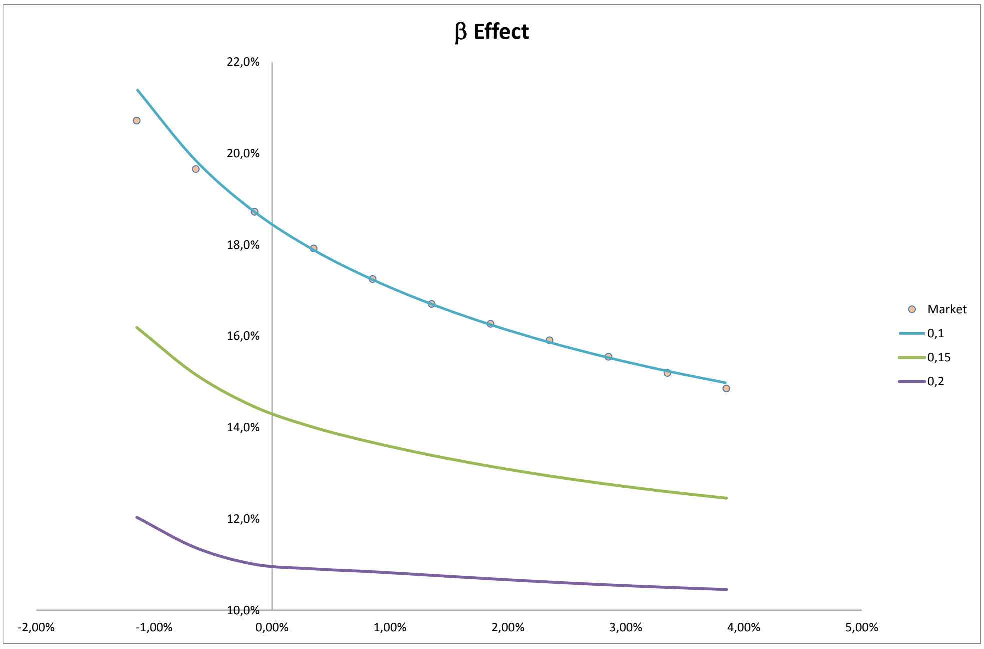


Figure 6.14: Beta effect for input in Table 6.3

	Approximation	Integral
Gauss-Legendre	3s	30s
Romberg	5min	> 1h

Table 6.6: The scale of the time required to calibrate one smile for the approximated and integral versions of function G 6.4

The price under The Free Boundary and the Normal SABR models is presented as a double integral, and numerical integration can be very time consuming. In order to circumvent this issue, Antonov presented an approximation of function G with a closed formula [11], thus reducing the pricing to a single integration. Table 6.6 gives the scale of the execution time of one smile for the integral version of G and for its approximation, when using either Guass-Legendre or Romberg integration methods.

From these results we can see that using Gauss-Legendre quadrature is much faster than Romberg integration. And if we calibrate 513 smiles using this method, it would take around 8 minutes with the approximation of G and around 4 hours with the integral version. Therefore the only "acceptable" setup is to use the approximated version combined with Gauss-Legendre integration.

One can note that each smile is calibrated independently from the others, thus it is possible to run the cube calibration on multithreading. This should significantly improve the swaption cube calibration time.

6.5 Greeks under the Mixture SABR model

Swaption prices obtained using the Shifted and Mixture SABR calibrated models are very close, the greeks however might be quite different. This means that the hedging is likely to be different depending on which model we chose to adopt, furthermore one of the key arguments of preferring the Mixture over the Shifted is its ability to hedge near-the-boundary options.

In this section we derive the standard greeks under the Mixture SABR based on our parametrization and calibration method. These expressions are derived for call prices under the following framework:

$$V^M = p\, V^F(F_0, \alpha_1, \beta_1, \gamma_1, 0) + (1-p)\, V^N(F_0, \alpha_2, 0, \gamma_2, \rho_2)$$

where

$$\alpha_1 = \alpha_1(\sigma_{\text{ATM}}, F_0) \qquad \alpha_2 = \alpha_1 |F_0|^{\beta_1}$$

$$\gamma_2 = \frac{\gamma_1}{1-\beta_1} \qquad p = \frac{\beta_1}{\beta_1 + |\rho_2|}$$

In this parametrization α_1 depends on both the ATM volatility and the forward rate, this is due to the fact that its value is fixed by fitting the ATM.

We have mentioned in the previous section that it is possible to calibrate this model with a greater number of parameters, we however argued that this would make the procedure very time consuming and not practical. At this point we can see another reason for not taking that path, indeed if we choose to calibrate the model without fixing the parameters α_2 and γ_2 the problem of properly defining the vega and volga risks arises. In a model where we have two initial volatilities (vol-of-vol) how do we define the vega (volga)?

Therefore if we want to hedge these risks under the Mixture SABR we need to link the initial volatility and the vol-of-vol of the zero-correlation Free Boundary and the Normal SABR. One way of doing so is the suggested parametrization.

Delta

Differentiating V^M with respect to F_0 yields:

$$\Delta = \frac{\partial V^M}{\partial F_0} + p\,\frac{\partial \alpha_1}{\partial F_0}\,\frac{\partial V^{ZC}}{\partial \alpha_1} + (1-p)\,\frac{\partial \alpha_2}{\partial F_0}\,\frac{\partial V^N}{\partial \alpha_2} \tag{6.18}$$

In practice we compute the greeks using finite differences. We should note that in order to compute $\frac{\partial \alpha_1}{\partial F_0}$ using finite differences, we write:

$$\frac{\partial \alpha_1}{\partial F_0} = \frac{\alpha_1(F_0 + \delta F_0, \sigma_{\text{ATM}}) - \alpha_1(F_0 - \delta F_0, \sigma_{\text{ATM}})}{2\,\delta F_0}$$

where $\alpha_1(F_0 + \delta F_0, \sigma_{\text{ATM}})$ and $\alpha_1(F_0 - \delta F_0, \sigma_{\text{ATM}})$ are computed by fitting the ATM volatility given that the forward rate is respectively $F_0 + \delta F_0$ and $F_0 - \delta F_0$. The derivative of α_2 with respect to F_0 is obtained similarly as we only need to multiply α_1 by $|F_0 \pm \delta F_0|$

Vega

As the initial volatilities are linked, we define the vega as the derivative of the price with respect to α_1.

$$\begin{aligned} \text{vega} &= p\,\frac{\partial V^{ZC}}{\partial \alpha_1} + (1-p)\,\frac{\partial \alpha_2}{\partial \alpha_1}\,\frac{\partial V^N}{\partial \alpha_2} \\ &= p\,\frac{\partial V^{ZC}}{\partial \alpha_1} + (1-p)\,|F_0|^{\beta_1}\frac{\partial V^N}{\partial \alpha_2} \end{aligned} \tag{6.19}$$

This expression of vega does not take into account the correlation between the forward rate and the volatility under the Normal SABR. This contribution was studied by Bartlett in [12] for the general SABR model, and it was shown that another term should be added. For the Normal SABR, this the additional term is:

$$\frac{\rho_2}{\gamma_2}\frac{\partial \alpha_2}{\partial F_0^{(2)}}\frac{\partial V^N}{\partial \alpha_2}$$

Unfortunately this expression depends on $F_0^{(2)}$ and not F_0. And since we don't have direct control over $F_0^{(2)}$, we cannot use it. This being said, it is worth mentioning that the contribution of this term in the classical SABR model is minimized when one hedges both vega and delta risks.

Volga

As the vol-of-vols are linked, we define the volga as the derivative of the price with respect to γ_1:

$$\begin{aligned} \text{volga} &= p\,\frac{\partial V^F}{\partial \gamma_1} + (1-p)\,\frac{\partial \gamma_2}{\partial \gamma_1}\,\frac{\partial V^N}{\partial \gamma_2} \\ &= p\,\frac{\partial V^F}{\partial \gamma_1} + (1-p)\,\frac{1}{1-\beta_1}\,\frac{\partial V^N}{\partial \gamma_2} \end{aligned} \tag{6.20}$$

Chapter 7

Conclusion

Each of the SABR model extensions to negative rates, has its advantages and drawbacks. First we have seen that shifting the SABR model produces excellent results in terms of fitting market data, but requires to set the shift prior to calibration.

Then we have studied the Free Boundary SABR [8]. This model represents an elegant and natural approach for taking into account the possibility of negative rates without any prior adjustments. Unfortunately it also has its drawbacks. Namely, despite the fact that the zero correlation formula is exact, the mapping technique allowing the parametrization for the general correlation case does not work for all possible parameter combinations. Sometimes this parametrization leads to negative initial volatility. In particular this mapping instability can be very problematic in calibration.

The Mixture SABR Model [2] offers a stable solution to circumvent this parametrization issues. It has proven to be a robust model, that can be efficiently calibrated. However some issues have risen, namely the incapacity of this model to fit market prices for extreme negative strikes. Nevertheless the quality of this quotes are questionable as these might actually represent extrapolated prices by a calibrated shifted SABR model.

Appendix A: Approximation of G

The prices of both the zero-correlation Free boundary and Normal SABR are presented as double integrals. Under this form, these models are not suitable for calibration as the process can be very time consuming. In order to deal with this, Antonov presented an approximated version of the inner integral (kernel function G), thus reducing the pricing formulas to a single integral.

The integral version of G is given by:

$$G(t,s) = 2\sqrt{2}\frac{e^{-\frac{t}{8}}}{t\sqrt{2\pi t}}\int_s^\infty u\, e^{-\frac{u^2}{2t}}\sqrt{\cosh u - \cosh s}\, du$$

The approximated version was presented in [11]. It was stated that the function G can be closely approximated by:

$$G(t,s) \simeq \sqrt{\frac{\sinh s}{s}}e^{-\frac{s^2}{2t}-\frac{t}{8}}\big(R(t,s)+\delta R(t)\big) \tag{1}$$

where

$$\begin{aligned}
R(t,s) =& 1 + \frac{3tg(s)}{8s^2} - \frac{5t^2(-8s^2+3g^2(s)+24g(s))}{128s^4} \\
&+ \frac{35t^3(-40s^2+3g^3(s)+24g^2(s)+120g(s))}{1024s^6} \\
g(s) =& s\coth(s) - 1 \\
\delta R(t) =& e^{\frac{t}{8}} - \frac{3072+384t+24t^2+t^3}{3072}
\end{aligned}$$

The correction δR is defined such that $G(t,0) = 1$. This approximation is derived using a small-time expansion based on the McKean Kernel result from [1].

In general this formula gives excellent results. However when s small it degenerates and yield inaccurate values. In order to deal with this, Antonov suggested to use its fourth-order approximation for small s. In this appendix, we derive this expansion.

Let's start by expanding the function $d : s \to \sqrt{\frac{\sinh s}{s}}$. We have

$$\begin{aligned}
&\sinh s = s + \frac{s^3}{3!} + \frac{s^5}{5!} + o(s^5) \\
&\implies \frac{\sinh s}{s} = 1 + \frac{s^2}{3!} + \frac{s^4}{5!} + o(s^4) \\
&\implies \sqrt{\frac{\sinh s}{s}} = (1 + \frac{s^2}{3!} + \frac{s^4}{5!} + o(s^4))^{\frac{1}{2}} \\
&\qquad\qquad = 1 + \frac{s^2}{12} + \frac{7}{1440}s^4 + o(s^4)
\end{aligned}$$

Thus

$$\sqrt{\frac{\sinh s}{s}} = 1 + \frac{s^2}{12} + \frac{7}{1440}s^4 + o(s^4) \tag{2}$$

In order to expand R, we need to expand each numerator to a different order. Let N_1, N_2 and N_3 denote the numerators of R:

$$N_1(t,s) = 3tg(s)$$
$$N_2(t,s) = 5t^2(-8s^2 + 3g^2(s) + 24g(s))$$
$$N_3(t,s) = 35t^3(-40s^2 + 3g^3(s) + 24g^2(s) + 120g(s))$$

N_1 should be expanded to the sixth-order, N_2 to the eight-order and N_3 to the tenth-order. This implies that we need to expand g to the tenth-order. One can show that:

$$g(s) = \frac{1}{3}s^2 - \frac{1}{45}s^4 + \frac{2}{945}s^6 - \frac{1}{4725}s^8 + \frac{2}{93555}s^{10} + o(s^{10})$$
$$g^2(s) = \frac{1}{9}s^4 - \frac{2}{135}s^6 + \frac{1}{525}s^8 - \frac{2}{8505}s^{10} + o(s^{10})$$
$$g^3(s) = \frac{1}{27}s^6 - \frac{1}{135}s^8 + \frac{17}{14175}s^{10} + o(s^{10})$$

From these expressions we find:

$$N_1(t,s) = 3t(\frac{1}{3}s^2 - \frac{1}{45}s^4 + \frac{2}{945}s^6) + o(s^6)$$
$$N_2(t,s) = 5t^2(-\frac{1}{5}s^4 + \frac{2}{315}s^6 + \frac{1}{1575}s^8) + o(s^8)$$
$$N_3(t,s) = 35t^3(\frac{s^6}{105} + \frac{67}{1575}s^8 + \frac{1}{1925}s^{10}) + o(s^{10})$$

Therefore R small s expansion is given by:

$$\begin{aligned} R(t,s) =& 1 + \frac{t}{8}(1 - \frac{1}{15}s^2 + \frac{2}{315}s^4) - \frac{t^2}{128}(-1 + \frac{2}{63}s^2 + \frac{1}{315}s^4) \\ &+ \frac{35t^3}{1024}(\frac{1}{105} + \frac{67}{1575}s^2 + \frac{1}{1925}s^4) + o(s^4) \end{aligned} \tag{3}$$

We note that this expression verifies $G(t,0) = 1$. Using the small-time expansion 1 with these small s expansions (when appropriate) yields excellent results. Adopting this approach reduces significantly the pricing time execution.

Bibliography

[1] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing Smile Risk. *Wilmott Magazine*, pages 84–108, July 2002.

[2] A. Antonov, M. Konikov, and M. Spector. Mixing SABR models for Negative Rates. *SSRN paper*, August 2015.

[3] Marc Henrard. *Interest Rate Modelling in the Multi-Curve Framework: Foundations, Evolution and Implementation*. 2014.

[4] John C. Hull. *Options, Futures and Other Derivatives (8th Edition)*.

[5] Jan Obloj. Fine-Tune your smile. Correction to Hagan et al. *Imperial college, working paper*, 2008.

[6] Louis Paulot. Asymptotic Implied Volatility at the Second Order With Application to the SABR Model. *working paper, Sophis Technology*, 2009.

[7] A. Antonov and M. Spector. Advanced analytics for the SABR model. *SSRN paper*, March 2012.

[8] A. Antonov, M. Konikov, and M. Spector. The Free Boundary SABR: Natural Extension to Negative Rates. *SSRN paper*, January 2015.

[9] A. Antonov, M. Konikov, D. Rufino, and M. Spector. Exact Solution to CEV Model with uncorrelated Stochastic Volatility. *SSRN paper*, 2014.

[10] D.Breeden and R. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, pages 621–651, 1978.

[11] A. Antonov, M. Konikov, and M. Spector. SABR spreads its wings. *Risk Magazine*, pages 58–63, August 2013.

[12] B. Bartlett. Hedging under SABR Model. *WILMOTT magazine*, pages 2–6, July/August 2006.

[13] H. Berestycki, J. Busca, and I. Florent. Computing the Implied Volatility in Stochastic Volatility Models. *Computational Finance 9:1*, 2004.