Explicit SABR Calibration through Simple Expansions

Fabien Le Floc'h[⋆] and Gary Kennedy[†]

[⋆]*Calypso Technology, 106 rue de La Boétie, 75008 Paris*
[†]*Clarus Financial Technology, London*

ABSTRACT   *The SABR stochastic volatility model is a very popular interpolator of implied volatilities, with a given dynamic. This paper presents a simple and very fast method to calibrate the SABR model to given market volatilities, that is to imply the SABR parameters from a given market smile.*

KEY WORDS: stochastic volatility, SABR, calibration, implied volatility, finance

1.   Introduction

The SABR stochastic volatility model (Hagan *et al.*, 2002) enjoys a high popularity as interpolator of implied volatilities, because, in practice, it fits rather well market implied volatility smiles for interest rate derivatives, fx derivatives and equity derivatives with a few parameters α, ρ, ν, and the dynamic of it can be easily controlled through its β parameter, usually defined from historical series analysis for the relevant market.

There are however known shortcomings: it is not arbitrage free as the probability density can become negative for low strikes and long maturities. Many authors have proposed various improvements to the original formula (Oblój, 2008; Johnson and Nonas, 2009; Paulot, 2009; Benaim *et al.*, 2008). More recently the focus has been on finite difference techniques to guarantee the arbitrage-free property (Andreasen and Huge, 2011; Hagan *et al.*, 2014; Le Floc'h and Kennedy, 2014) in a shifted SABR framework, to allow for negative rates.

In practice, the adjustments to the original formula do not matter much in order to find a good initial guess for the calibration procedure. Once a good initial guess is found, a fast local minimizer like Levenberg-Marquardt can be used to calibrate the specific SABR implementation.

A first calibration method was described in (West, 2005), fitting α exactly to the at-the-money implied volatility, and reducing the problem to a two-dimensional minimization in (ρ, ν). The Nelder-Mead method is proposed as the minimizer. However, it still requires an initial guess, and can sometimes be unstable, especially with constraints (Le Floc'h, 2012). A numerical method to find an initial guess that fits the at-the-money volatility exactly and the at-the-money skew is proposed in (Gauthier and Rivaille, 2009). It requires the numerical solution of a two-dimensional non linear system of equations.

In contrast, the method proposed here is simple and explicit; a system of equations to fit the at-the-money volatility, skew, and curvature is solved analytically. A similar approach was

2   *Fabien Le Floc'h and Gary Kennedy*

applied to the Heston model in (Forde *et al.*, 2012) with the difference that, in their case, the Heston parameters describe a full volatility surface and the initial guess procedure relies on two expiries plus the short time at-the-money volatility to solve the 5 Heston parameters exactly. In the SABR case, there is just one expiry to consider and 3 parameters to fit. We will first present our method for the lognormal formula as well as for the normal formula, and then show its accuracy in calibrating real world smiles on the arbitrage-free model from Hagan *et al.* (2014) compared to a global optimization via differential evolution.

2.   Classic SABR normal and lognormal formulas

2.1   *The SABR model*

In the shifted SABR model, the asset forward follows the following stochastic differential equation:

$$dF = V(F+b)^\beta dW_1 \tag{1}$$

$$dV = \nu V dW_2 \tag{2}$$

with W_1, W_2 Brownian motions correlated with correlation ρ, and $F(0) = f$, $V(0) = \alpha$. ν represents the volatility of volatility, b is a displacement allowing for negative rates ($b = 0$ for the classic SABR model).

2.2   *Normal formula*

Given the low rates environment we currently experience, it is now common practice to quote swaptions in terms of normal volatility (bpvol). Taking into account the remarks from Oblój (2008) for the case of the lognormal formula to ensure consistency when $\beta \to 1$, the expansion of the normal volatility adapted from Hagan *et al.* (2002) for the shifted SABR model is:

for $f \neq K$ and $\beta \in [0, 1]$

$$\sigma_N(K) = \frac{f-K}{x(\zeta(K))} \left[1 + \left(g(K) + \frac{1}{4}\rho\nu\alpha\beta(f+b)^{\frac{\beta-1}{2}}(K+b)^{\frac{\beta-1}{2}} + \frac{1}{24}(2-3\rho^2)\nu^2 \right) T \right] \tag{3}$$

with

$$g(K) = \frac{1}{24}(\beta^2 - 2\beta)(f+b)^{\beta-1}(K+b)^{\beta-1}\alpha^2 \tag{4}$$

$$\zeta(K) = \frac{\nu}{\alpha(1-\beta)} \left((f+b)^{1-\beta} - (K+b)^{1-\beta} \right) \tag{5}$$

$$x(\zeta) = \frac{1}{\nu} \log \left(\frac{\sqrt{1 - 2\rho\zeta + \zeta^2} - \rho + \zeta}{1 - \rho} \right) \tag{6}$$

When $\beta = 1$,

$$g(K) = -\frac{1}{24}\alpha^2 \ , \ \zeta(K) = \frac{\nu}{\alpha} \log \left(\frac{f+b}{K+b} \right)$$

When $\beta = 0$,

$$g(K) = 0 \ , \ \zeta(K) = \frac{\nu}{\alpha}(f-K)$$

When $\nu \to 0$,

$$x(K) \sim \frac{1}{\nu}\zeta(K)$$

When $f = K$,

$$\sigma_N(f) = \alpha(f + b)^\beta \left[1 + \left(g(f) + \frac{1}{4}\rho\nu\alpha\beta(f + b)^{\beta-1} + \frac{1}{24}(2 - 3\rho^2)\nu^2 \right) T \right] \tag{7}$$

2.3   Lognormal formula

The lognormal formula is very similar to the normal formula, $f - K$ becomes $\log(\frac{f}{K})$ and g includes a small adjustment:

for $f \neq K$ and $\beta \in [0, 1]$,

$$\sigma_B(K) = \frac{1}{x(\zeta(K))} \log\left(\frac{f + b}{K + b} \right) \left[1 + \left(g(K) + \frac{1}{4}\rho\nu\alpha\beta(f + b)^{\frac{\beta-1}{2}}(K + b)^{\frac{\beta-1}{2}} + \frac{1}{24}(2 - 3\rho^2)\nu^2 \right) T \right] \tag{8}$$

with

$$g(K) = \frac{1}{24}(\beta - 1)^2(f + b)^{\beta-1}(K + b)^{\beta-1}\alpha^2 \tag{9}$$

and $\zeta(K), x(\zeta)$ as in equations (5) and (6).

When $\beta = 1$, $g(K) = 0$.

When $\beta = 0$, $g(K) = \frac{\alpha^2}{24(f+b)(K+b)}$.

When $f = K$,

$$\sigma_B(f) = \alpha(f + b)^{\beta-1} \left[1 + \left(g(f) + \frac{1}{4}\rho\nu\alpha\beta(f + b)^{\beta-1} + \frac{1}{24}(2 - 3\rho^2)\nu^2 \right) T \right] \tag{10}$$

It is important not to forget that Hagan derived the lognormal formula as an approximation of the normal formula (Hagan *et al.*, 2002): it won't be accurate or realistic at all in cases where the implied volatility is very high (see Table 1). This is not so realistic for market implied volatilities, but it can have an influence in the global minimization procedure, as global minimizers like differential evolution will try out extreme values.

Of course, in order to obtain a lognormal volatility, it is also possible to solve, with high accuracy, for the Black implied volatility of a given normal volatility: we can firstly compute corresponding Bachelier option price with a fast evaluation method (Le Floc'h, 2014) and secondly use a robust and accurate Black implied volatility solver (Jäckel, 2013; Li and Lee, 2011).

Table 1.: Black volatility obtained from the direct lognormal approximation compared the one solved from the normal approximation using
$\alpha = 3.24, \beta = 1.0, \rho = -0.998, \nu = 1.69, f = 2014, K = f, T = 0.48$.

Formula	Black volatility	Option price
Lognormal	0.9325	510.19
Normal	0.2526	140.41

For performance reasons, we prefer to rely on the normal formula if the calibration is for bpvols and the lognormal formula if the calibration is for Black volatilities.

3.   Andersen & Brotherton-Ratcliffe expansions

As explained in (Le Floc'h and Kennedy, 2014), the arbitrage-free SABR method from Hagan *et al.* (2014) is equivalent to a local volatility expansion where the vanilla call option prices

4   *Fabien Le Floc'h and Gary Kennedy*

follow the normal forward Dupire PDE:

$$\frac{\partial V_{call}}{\partial T}(T, K) = \frac{1}{2}\vartheta^2(T, K)\frac{\partial^2 V_{call}}{\partial K^2}(T, K) \tag{11}$$

with initial condition $V_{call}(0, K) = (f - K)^+$ and

$$\vartheta(T, K) = D(K)E(T, K),\ E(T, K) = e^{\frac{1}{2}\rho\nu\alpha\Gamma(K)T},\ \Gamma(K) = \frac{(K + b)^\beta - (f + b)^\beta}{K - f} \tag{12}$$

$$D(K) = \sqrt{\alpha^2 + 2\alpha\rho\nu y(K) + \nu^2 y(K)^2}K^\beta,\ y(K) = \frac{(K + b)^{1-\beta} - (f + b)^{1-\beta}}{1 - \beta} \tag{13}$$

Following the same idea as in (Karlsmark, 2013), normal and lognormal implied volatility expansions can be obtained from the local volatility using the technique of (Andersen and Brotherton-Ratcliffe, 2005). The first step is to remove the time dependency by a change of variable:

$$u(T) = \int_0^T E^2(t, K)dt = \frac{e^{\rho\nu\alpha\Gamma(K)T} - 1}{\rho\nu\alpha\Gamma(K)} \tag{14}$$

The PDE becomes:

$$\frac{\partial V_{call}}{\partial u}(u, K) = \frac{1}{2}D^2(K)\frac{\partial^2 V_{call}}{\partial K^2}(u, K) \tag{15}$$

3.1   A new lognormal formula

Applying Andersen and Brotherton-Ratcliffe (2005, proposition B.1) to equation (15) gives the following lognormal expansion of the volatility:

$$\Omega(u, K) = \Omega_0(K) + \Omega_1(K)u + \mathcal{O}(u^2) \tag{16}$$

with

$$\Omega_0(K) = \frac{\log\left(\frac{f+b}{K+b}\right)}{\int_K^f D^{-1}(k)dk} \tag{17}$$

$$\Omega_1(K) = -\frac{\Omega_0(K)}{\left(\int_K^f D^{-1}(k)dk\right)^2}\log\left(\Omega_0(K)\sqrt{\frac{(f + b)(K + b)}{D(f)D(K)}}\right) \tag{18}$$

Those terms can be computed analytically as

$$\int_K^f D^{-1}(k)dk = \frac{1}{\nu}\log\left(\frac{\sqrt{1 + 2\rho\frac{\nu}{\alpha}y(K) + \frac{\nu^2}{\alpha^2}y^2(K)} - \rho - \frac{\nu}{\alpha}y(K)}{1 - \rho}\right) \tag{19}$$

Let's define $\zeta(K) = -\frac{\nu}{\alpha}y(K) = \frac{\nu}{\alpha(1-\beta)}\left((f + b)^{1-\beta} - (K + b)^{1-\beta}\right)$, that is the same ζ as in the classic lognormal formula (Equation 5, we have:

$$\int_K^f D^{-1}(k)dk = x(\zeta(K)) \tag{20}$$

where x is defined as in the classic lognormal formula (Equation 6). We go back to the variable T using $\sigma(K, T)^2 T = \Omega(K, u(T))^2 u(T)$, which leads to the following lognormal implied

volatility expansion:

$$\sigma(K,T) = \frac{\Omega_0(K)u^{\frac{1}{2}}(T) + \Omega_1(K)u^{\frac{3}{2}}(T)}{\sqrt{T}} \tag{21}$$

As the approximation is only order-1 in time, the exponential term can be expanded as well:

$$u(T) = T + \frac{1}{2}\rho\nu\alpha\Gamma(K)T^2 + \mathcal{O}(T^3) \tag{22}$$

This finally gives the expansion σ_B^{AB}:

$$\sigma_B^{AB}(K) = \Omega_0(K)\left(1 + \frac{1}{4}\rho\nu\alpha\Gamma(K)T\right) + \Omega_1(K)T + \mathcal{O}(T^2) \tag{23}$$

or

$$\sigma_B^{AB}(K) = \frac{1}{x(\zeta(K))}\log\left(\frac{f+b}{K+b}\right)\left[1 + \left(g(K) + \frac{1}{4}\rho\nu\alpha\Gamma(K)\right)T\right] + \mathcal{O}(T^2) \tag{24}$$

with

$$g(K) = -\frac{1}{x^2(\zeta(K))}\log\left(\frac{\log\left(\frac{f+b}{K+b}\right)}{x(\zeta(K))}\sqrt{\frac{(f+b)(K+b)}{D(f)D(K)}}\right)$$

The later formula is particularly interesting since it matches the classic lognormal formula (Equation 8) as well as the formulas found in (Hagan *et al.*, 2002; Oblój, 2008) exactly at-the-money while being very close to the arbitrage free PDE solution in practice (see Table 2, Figure 1).

Table 2.: Prices and implied volatilities with the various expansions using parameters of (Hagan, 2014) $\alpha = 0.35, \beta = 0.25, \rho = 0.25, \nu = 1, T = 2, f = 1$. "classic" denotes the formula from Equation (8) and "new" corresponds to Equation (24)

Strike	Method	Price	Black vol
1.0	PDE	0.2246	40.35%
1.0	classic	0.2274	40.87%
1.0	new	0.2274	40.87%
1.5	PDE	0.1092	42.91%
1.5	classic	0.1265	46.15%
1.5	new	0.1088	42.85%

3.2   *A new normal formula*

Applying Andersen and Brotherton-Ratcliffe (2005, proposition B.2) to equation (15) gives the following normal expansion of the volatility:

$$\Omega(u,K) = \Omega_0(K) + \Omega_1(K)u + \mathcal{O}(u^2) \tag{25}$$

with

$$\Omega_0(K) = \frac{f-K}{\int_K^f D^{-1}(k)dk} \tag{26}$$

$$\Omega_1(K) = -\frac{\Omega_0(K)}{\left(\int_K^f D^{-1}(k)dk\right)^2}\log\left(\frac{\Omega_0(K)}{\sqrt{D(f)D(K)}}\right) \tag{27}$$

6   *Fabien Le Floc'h and Gary Kennedy*

This results in the normal volatility expansion:

$$\sigma(K, T) = \frac{\Omega_0(K)u^{\frac{1}{2}}(T) + \Omega_1(K)u^{\frac{3}{2}}(T)}{\sqrt{T}} \tag{28}$$

Expanding the exponential term, we finally have:

$$\sigma_N^{AB}(K) = \frac{f - K}{x(\zeta(K))} \left[1 + \left(g(K) + \frac{1}{4}\rho\nu\alpha\Gamma(K)\right)T\right] + \mathcal{O}(T^2) \tag{29}$$
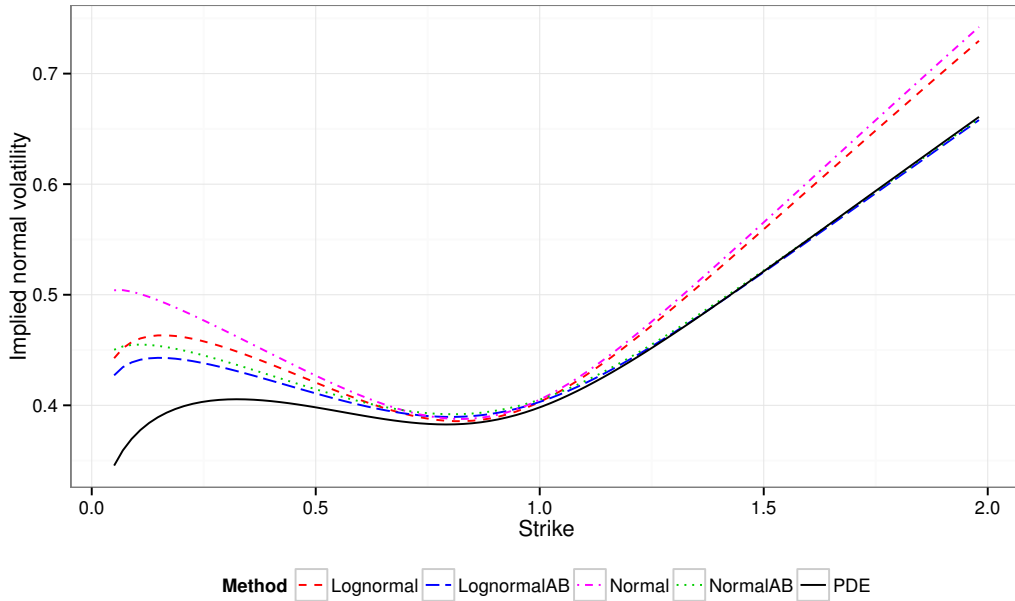
with

$$g(K) = -\frac{1}{x^2(\zeta(K))} \log\left(\frac{f - K}{x(\zeta(K))\sqrt{D(f)D(K)}}\right)$$

and $\zeta(K), x(\zeta)$ defined as in equations (5) and (6).

Again, σ_N^{AB} matches the classic normal formula (Equation 3) as well as the formula from (Hagan *et al.*, 2002) exactly at the money.

Figure 1.: Implied normal volatilities using parameters of (Hagan, 2014)
$\alpha = 0.35, \beta = 0.25, \rho = 0.25, \nu = 1, T = 2, f = 1.$

4.   Explicit initial guess

4.1   Lognormal volatility

The idea is to find α, ρ, ν that matches the at-the-money implied volatility σ_0, skew σ_0' and curvature σ_0''. A direct application of Hagan formula would lead to a system of trivariate polynomials of degrees 3 and 4, which would then require a numerical method. Instead, we rely a simple lower order expansion of the lognormal formula in the variable $z = \log\left(\frac{K+b}{f+b}\right)$ around $z = 0$ that still captures the smile around the at-the-money level accurately:

$$\sigma_B(z) = \alpha(f+b)^{\beta-1} + \frac{1}{2}\left(\rho\nu - (1-\beta)\alpha(f+b)^{\beta-1}\right)z$$

$$+ \frac{1}{12\alpha(f+b)^{\beta-1}}\left((1-\beta)^2(\alpha(f+b)^{\beta-1})^2 + \nu^2(2-3\rho^2)\right)z^2 \qquad (30)$$

This is the same expansion used in (Hagan *et al.*, 2002, equation (3.1a)) to explain the phenomenology and dynamic of the SABR model. It can also be found from the expansion of (Lorig *et al.*, 2014) by using the order-0 expansion for the at-the-money implied volatility, the order-1 expansion for the at-the-money skew, and the order-2 for the at-the-money curvature and ignore terms quadratic in time to expiry T^2. Or more simply it can be rederived from a Taylor expansion of Equation (8), neglecting higher order terms as in Appendix A. Note that the new lognormal formula (Equation 24) would also lead to the same expansion. We then have:

$$\begin{cases} \sigma_0 & = \alpha(f+b)^{\beta-1} \\ \sigma_0' & = \frac{1}{2}\left(\rho\nu - (1-\beta)\sigma_0\right) \\ \sigma_0'' & = \frac{1}{3\sigma_0}\nu^2 + \frac{1}{6\sigma_0}\left((1-\beta)^2\sigma_0^2 - 3\rho^2\nu^2\right) \end{cases} \qquad (31)$$

Note that the derivatives are expressed on log-moneyness, z.

This system can be exactly solved to give a first guess α_0, ρ_0, ν_0:

$$\begin{cases} \alpha_0 & = \sigma_0(f+b)^{1-\beta} \\ \nu_0^2 & = 3\sigma_0\sigma_0'' - \frac{1}{2}(1-\beta)^2\sigma_0^2 + \frac{3}{2}\left(2\sigma_0' + (1-\beta)\sigma_0\right)^2 \\ \rho_0 & = \frac{1}{\nu_0}\left(2\sigma_0' + (1-\beta)\sigma_0\right) \end{cases} \qquad (32)$$

If $\nu_0^2 < 0$, we force ρ_0 in $\{-1,1\}$ with $\rho_0 = \text{sgn}\left(2\sigma_0' + (1-\beta)\sigma_0\right)$, and recompute ν_0 ignoring the curvature through $\nu_0 = \frac{1}{\rho_0}\left(2\sigma_0' + (1-\beta)\sigma_0\right)$. In practice, this allows to find a good initial guess even when the optimal ρ is near -1 and when the smile has very little curvature. We then refine this first guess by solving exactly for the at-the-money volatility with the given ρ_0 and ν_0. α_1 is the root of the following third order polynomial [1]:

$$\frac{(1-\beta)^2 T}{24(f+b)^{2-2\beta}}\alpha^3 + \frac{\rho\beta\nu T}{4(f+b)^{1-\beta}}\alpha^2 + \left(1 + \frac{2-3\rho^2}{24}\nu^2 T\right)\alpha - \sigma_0(f+b)^{1-\beta} = 0 \qquad (33)$$

This is the same polynomial used by West (2005) to reduce the dimension of the calibration problem: α_1 is chosen as the smallest positive root. In theory, choosing α_1 closest to the previously calibrated α (either from the previous expiry, or from a previous day) could be better. In practice, it turns out to always be the smallest positive root. Our initial guess is then $(\alpha_1, \rho_0, \nu_0)$.

4.2   Normal volatility

When the input volatilities are normal (bpvol), we can just convert them to equivalent log-normal volatilities and apply the same method as in the previous section to find the initial guess. In our case, because the initial guess method depends mostly on volatilities closest to the money, we can also rely on another simple expansion based on the work of (Lorig *et al.*,

---

[1]The root can be determined using Cardano's formula

2014):

$$\sigma_B = v_0 - \frac{1}{2}v_0 z + \frac{1}{96}v_0(8z^2 + Tv_0^2(4 - Tv_0^2))$$

$$+ \frac{1}{192}Tzv_0^3(-12 + 5Tv_0^2) \tag{34}$$

with $v_0 = \frac{\sigma_N}{f}$ and $z = \log(\frac{K}{f})$, where the last term is of order-3.

Instead, we prefer to work directly with normal volatility and derive a simple expansion of the normal formula (Equation 3) in $z = \log\left(\frac{K+b}{f+b}\right)$ around $z = 0$ (see Appendix A);

$$\sigma_N(z) = \alpha(f+b)^\beta + \frac{1}{2}\left(\rho\nu(f+b) + \beta\alpha(f+b)^\beta\right)z$$

$$+ \left[\frac{1}{12\alpha(f+b)^{\beta-2}}(2\nu^2 - 3\rho^2\nu^2) + \frac{1}{4}\rho\nu(f+b) + \frac{1}{12}(\beta^2+\beta)\alpha(f+b)^\beta\right]z^2 \tag{35}$$

The new normal formula (Equation 29) would also lead to the same expansion.

Let $\sigma_0, \sigma_0', \sigma_0''$ be respectively the normal volatility, the slope and the curvature at the money expressed in log-moneyness z. We have:

$$\begin{cases} \sigma_0 & = \alpha(f+b)^\beta \\ \sigma_0' & = \frac{1}{2}\rho\nu(f+b) + \frac{1}{2}\beta\sigma_0 \\ \sigma_0'' & = \frac{(f+b)^2}{6\sigma_0}(2\nu^2 - 3\rho^2\nu^2) + \frac{1}{2}\rho\nu(f+b) + \frac{1}{6}(\beta^2+\beta)\sigma_0 \end{cases} \tag{36}$$

This system can be exactly solved to give a first guess α_0, ρ_0, ν_0:

$$\begin{cases} \alpha_0 & = \sigma_0(f+b)^{-\beta} \\ \nu_0^2 & = \frac{1}{(f+b)^2}\left[3\sigma_0\sigma_0'' - \frac{1}{2}(\beta^2+\beta)\sigma_0^2 - 3\sigma_0(\sigma_0' - \frac{1}{2}\beta\sigma_0) + \frac{3}{2}\left(2\sigma_0' - \beta\sigma_0\right)^2\right] \\ \rho_0 & = \frac{1}{\nu_0(f+b)}\left(2\sigma_0' - \beta\sigma_0\right) \end{cases} \tag{37}$$

As for the lognormal volatility guess, we refine this first guess by solving exactly for the at-the-money volatility with the given ρ_0 and ν_0. Then α_1 is a root of the following third order polynomial:

$$\frac{(\beta^2 - 2\beta)T}{24(f+b)^{2-2\beta}}\alpha^3 + \frac{\rho\beta\nu T}{4(f+b)^{1-\beta}}\alpha^2 + \left(1 + \frac{2 - 3\rho^2}{24}\nu^2 T\right)\alpha - \sigma_0(f+b)^{-\beta} = 0 \tag{38}$$

Note that this polynomial is slightly different from the lognormal volatility polynomial. Again we take the α_1 to be the smallest positive root.

4.3   *How to find the at-the-money volatility, slope and curvature?*

The simplest is to fit a parabola to the three closest points around the forward with coordinates $(z_{-1}, \hat\sigma_{-1}), (z_0, \hat\sigma_0), (z_1, \hat\sigma_1)$. This is equivalent to a 3 points finite difference on a non uniform grid. We then have:

$$\sigma_0 = z_0 z_1 w_{-1}\hat\sigma_{-1} + z_{-1}z_1 w_0\hat\sigma_0 + z_{-1}z_0 w_1\hat\sigma_1 \tag{39}$$

$$\sigma_0' = -(z_0 + z_1)w_{-1}\hat\sigma_{-1} - (z_{-1} + z_1)w_0\hat\sigma_0 - (z_{-1} + z_0)w_1\hat\sigma_1 \tag{40}$$

$$\sigma_0'' = 2w_{-1}\hat\sigma_{-1} + 2w_0\hat\sigma_0 + 2w_1\hat\sigma_1 \tag{41}$$

with

$$w_{-1} = \frac{1}{(z_{-1} - z_0)(z_{-1} - z_1)} \tag{42}$$

$$w_0 = \frac{1}{(z_0 - z_{-1})(z_0 - z_1)} \tag{43}$$

$$w_1 = \frac{1}{(z_1 - z_{-1})(z_1 - z_0)} \tag{44}$$

In our experience, using higher order approximations like a 5 points finite difference (equivalent to a quartic on 5 points), or a cubic spline does not lead to any visible improvement in accuracy on our problem.

When the input data is noisy, it can be useful to resort to a best fit parabola around the five or seven closest points from the forward, which can be reduced to solving simple linear system (Appendix B). When the volatility at the forward is already present in the input data, we can also make sure that the parabola passes exactly through it. We have noticed that including further away points in the estimation of the at-the-money curvature could be particularly important when the curvature is relatively small, for example for long-term options quotes (10 to 20 years) with a bit of noise. In some extreme situations, the three points curvature can be slightly negative and not allow for a good initial guess, as it can not represent such curvatures, while the seven points curvature will be accurate enough.

Another approach would be to repeat the initial guess procedure with a parabola on different sets of three points further away from the forward and select the guess corresponding to the best overall fit.

In our numerical tests we will select the guess that gives the minimum mean square error in volatilities between the two guesses stemming from the 3-points parabola and the 7-points parabola.

## 5. Almost exact inversion of SABR smiles

We use the SABR parameters found by calibration of the lognormal SABR model on the S&P500 options in December 2008 as starting point (Table 3). From those we regenerate a discrete set of implied volatilities for 12 strikes and 11 expiries using the lognormal SABR formula with $\beta = 1$. And finally we apply the explicit initial guess procedure to each expiry.

Table 3.: Reference SABR parameters with $f = 2016$ and explicit guess

Expiry	Reference parameters			Explicit guess			
years	α	ρ	ν	α	ρ	ν	vol RMSE
0.058	0.271	-0.345	1.010	0.271	-0.343	1.008	2.64e-04
0.153	0.256	-0.321	0.933	0.256	-0.319	0.934	1.36e-04
0.230	0.256	-0.346	0.820	0.256	-0.344	0.822	1.13e-04
0.479	0.255	-0.370	0.629	0.255	-0.369	0.631	8.25e-05
0.729	0.257	-0.403	0.528	0.257	-0.402	0.528	2.30e-05
1.227	0.260	-0.429	0.448	0.260	-0.429	0.447	6.15e-05
1.726	0.261	-0.440	0.392	0.261	-0.440	0.390	1.17e-04
2.244	0.262	-0.445	0.355	0.262	-0.445	0.352	1.59e-04
2.742	0.262	-0.445	0.329	0.262	-0.445	0.326	1.85e-04
3.241	0.262	-0.447	0.310	0.262	-0.447	0.306	2.13e-04
4.239	0.263	-0.452	0.284	0.263	-0.452	0.279	2.67e-04

The root mean square error in implied volatilities of the fit is always lower than $3 \cdot 10^{-4}$. ρ, ν are recovered with an accuracy higher than $5 \cdot 10^{-3}$ and α is recovered with and accuracy higher than 10^{-4}, independently of the expiry (see Table 3). The refinement α_1 allows to gain one order of magnitude in accuracy over α_0, we display the various stages of the initial guess procedure on the slice of expiry 0.479 on Figure 2.

Figure 2.: The various stages of the initial guess method for the slice of expiry 0.479 year
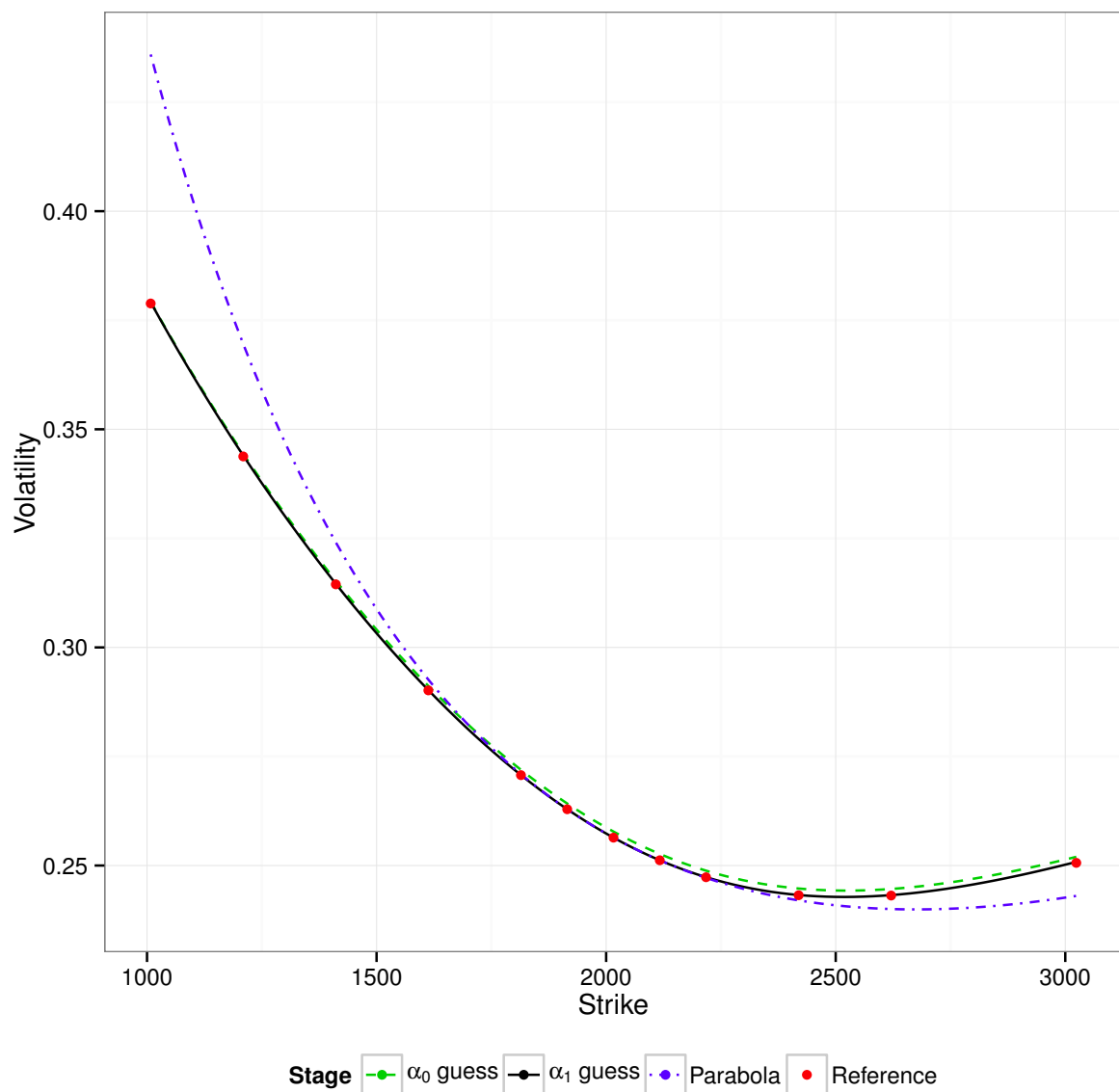
6.   Real world smiles

6.1   Calibration on the analytic formula

6.1.1   Equity option smiles

To calibrate SABR to equity implied volatility surfaces from December 2012, we minimize the weighted root mean square error in implied volatilities with a simple Gauss-Newton (GN) method adapted to handle rank-deficient systems (see Appendix C), placing more weight to volatilities around ATM +/- 20%. We compare the results obtained using either our explicit method as initial guess, or alternatively a guess found by running the differential evolution algorithm (Storn and Price, 1997) with a population size 20 on 1000 generations. Of course, instead of using the Gauss-Newton method, it is also possible to rely on the more robust Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963). In practice, however, our results would be exactly the same as the initial guess is always close enough so that the refinements of the Levenberg-Marquardt method are not needed. The resulting mean square error is displayed in Figure 3.

Figure 3.: Root mean square error of calibration per expiry on various equity surfaces
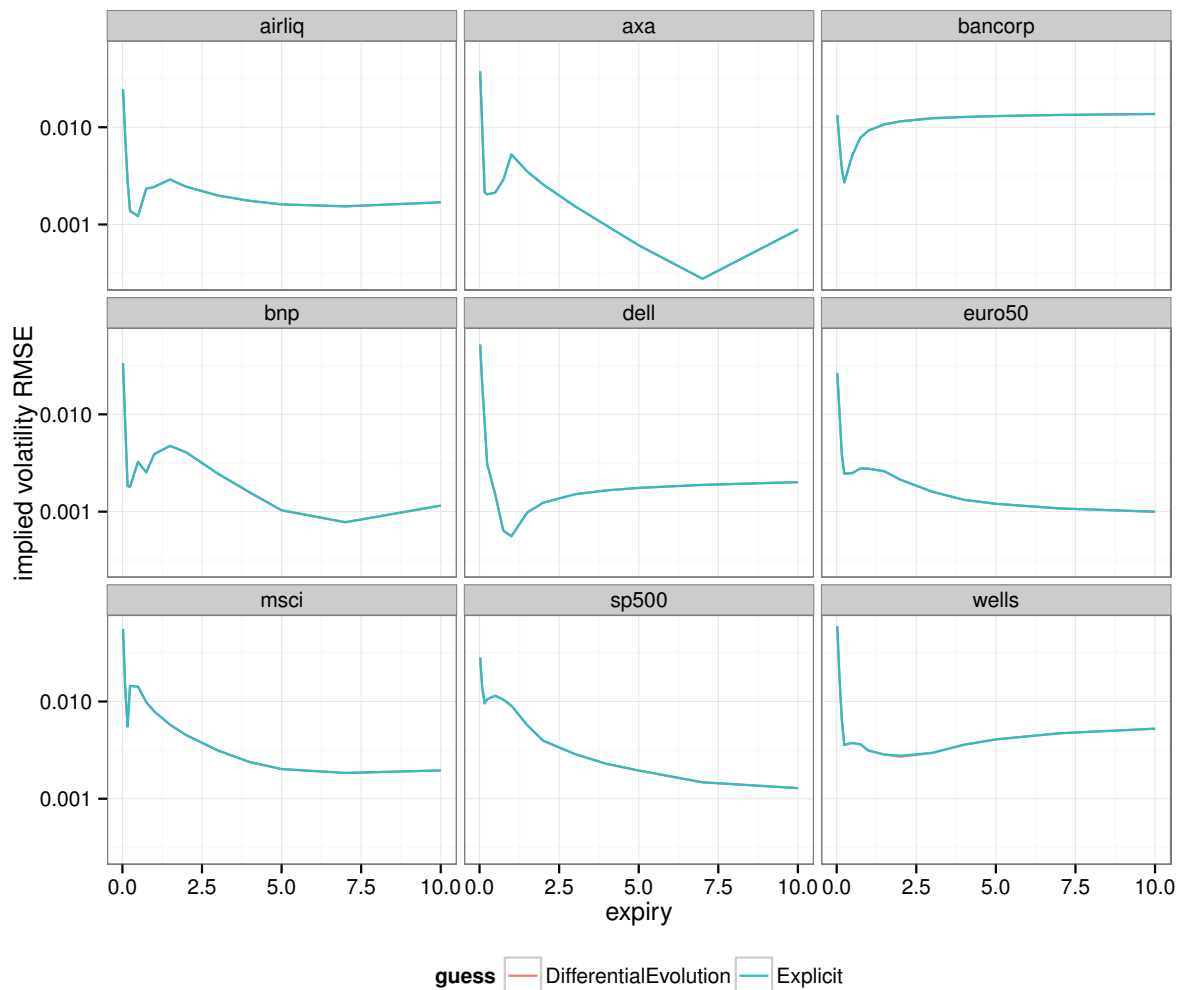
The minimization on the explicit guess leads to a mean square error in implied volatilities nearly indistinguishable from a minimization on the differential evolution guess. It is interesting however to look at the calibrated parameters values: the calibrated α shows some strong variations with the guess found by differential evolution, while it is relatively smooth with our explicit initial guess (Figure 4). ρ and ν behave similarly. We will take a closer look at why this happens in the next section.

Figure 4.: Calibrated α on various equity surfaces
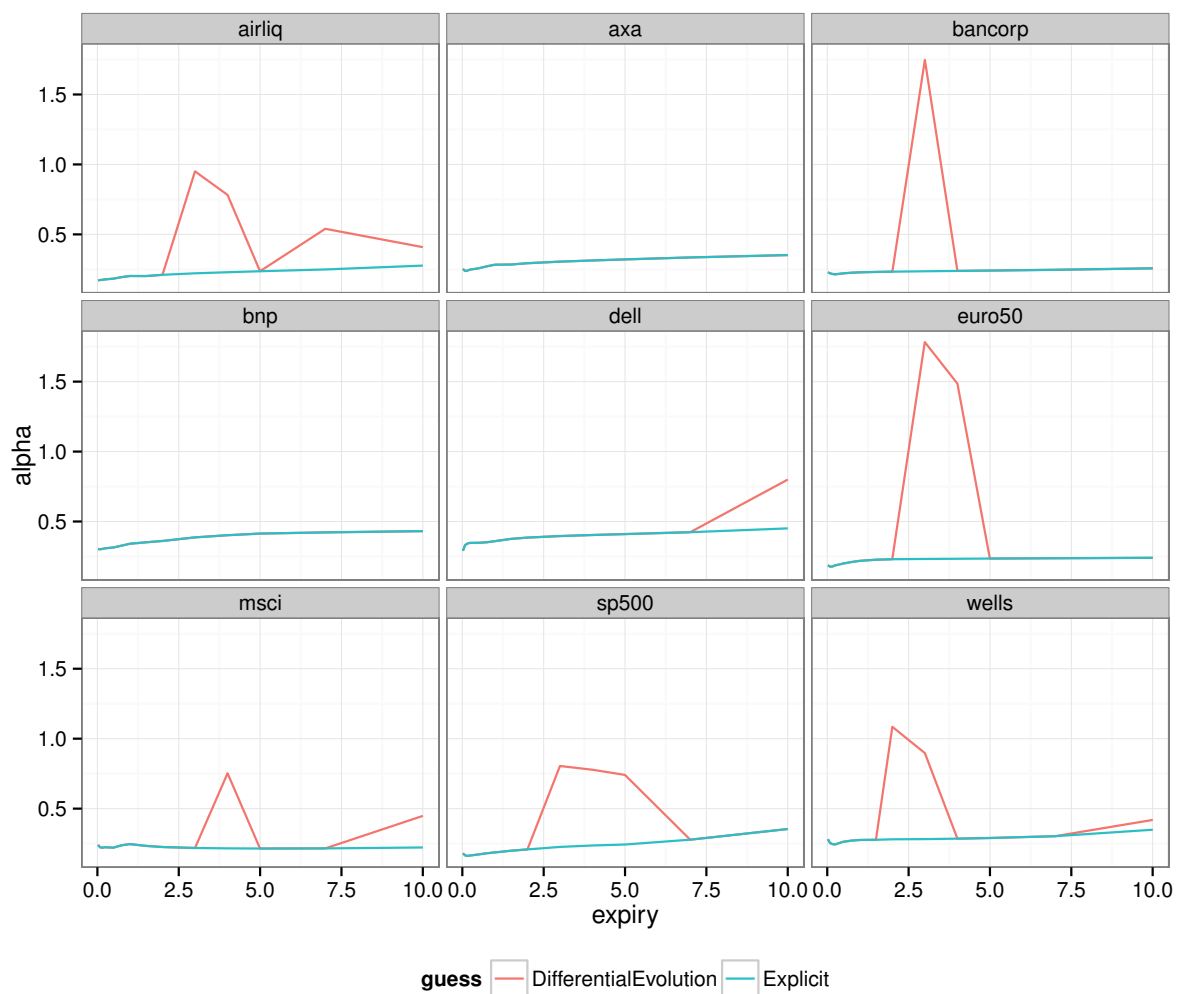
The initial guess is always fairly close to the true minimum: the median number of objective function evaluations in the Gauss-Newton method (corresponding to n SABR formula evaluations for n strikes on a given option expiry) is 3 per calibration (Figure 5). The number of objective function evaluations per calibration would be nearly the same using the Levenberg-Marquardt method instead of the Gauss-Newton method, except for a few expiries when the optimal ρ is close to -1 and the system almost degenerate, which makes then the Levenberg-Marquardt method slower to converge.

Figure 5.: Distribution of objective function evaluations per calibration with Gauss-Newton
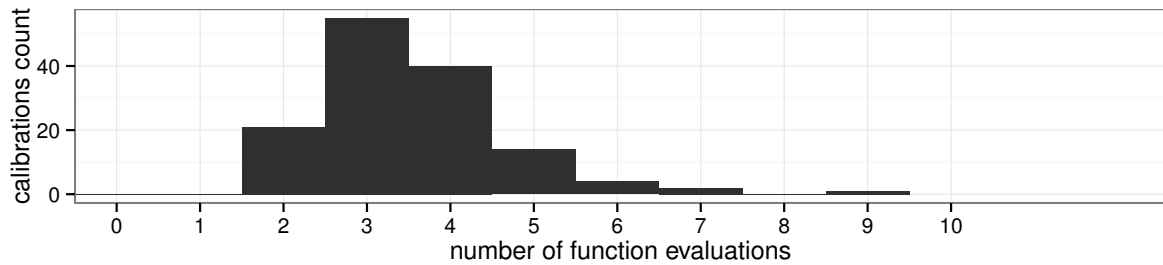
6.1.2   *Two SABRs for the same smile*

Figure 6 illustrates how two different sets of SABR parameters can lead to an equally good fit. We just picked the S&P 500 options expiring in 4 years used in the previous section. The

Figure 6.: Smiles corresponding to $\alpha = 0.237, \rho = -0.735, \nu = 0.398$ and $\alpha = 0.850, \rho = -0.742, \nu = 1.335$ with $\beta = 1, f = 1426.19, T = 4$
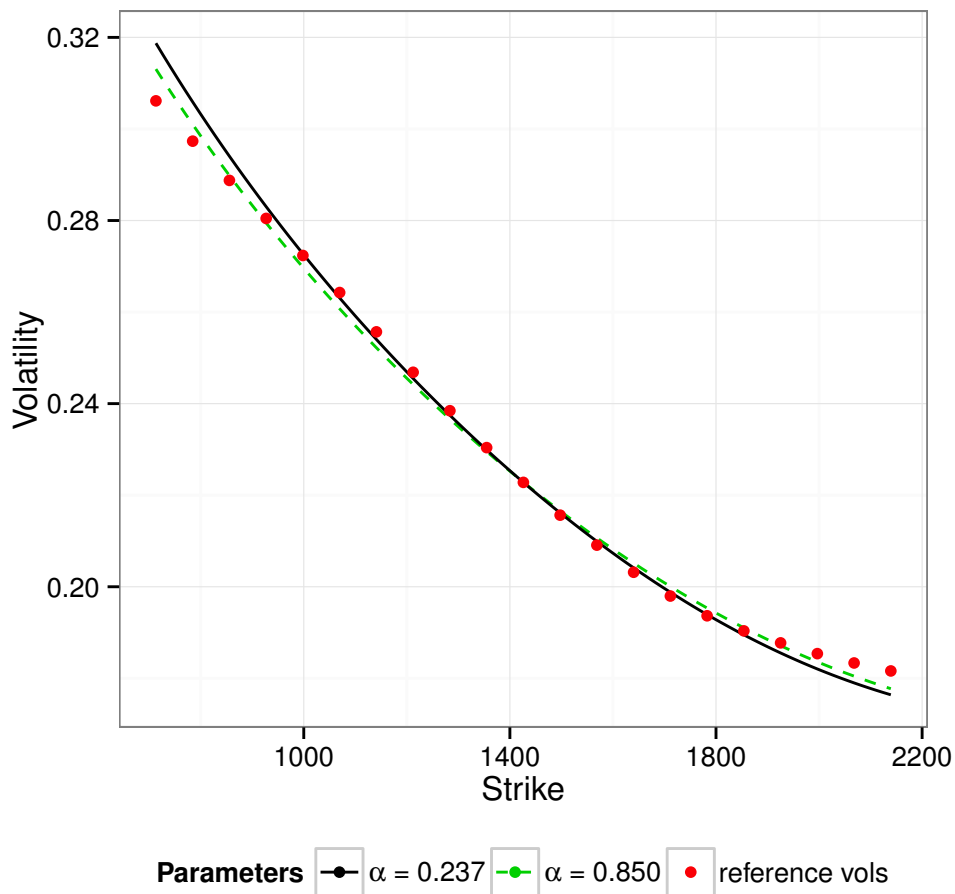
root mean square error is 0.00219269910750194 for the SABR parameters with lower α while

the one for the higher α is 0.002192699107502103, that is the lower α fits better by only 10^{-16}.

The fact that two different SABR parameters sets can lead to smiles with nearly the same error measure is the root of issues with differential evolution: it is not obvious which set of parameter is the right one without some additional constraint on the variation of the parameters. This can be resolved by some penalty term, but it is difficult in practice to find the correct penalty term that works for a variety of situations. A better approach would be look beyond the best candidate, and consider as alternative initial guesses the other candidates in the latest generation with very similar error measure and different (lower) α. In contrast, the explicit guess leads directly to the correct smile.

6.1.3   Swaption smiles

We apply the same calibration methodology as for the equity smiles cases, but relying on the normal formula with $\beta = 0.5$ instead (Equation 3). The resulting calibrated parameters are indistinguishable between the minimization on the explicit initial guess and the minimization on the guess found by differential evolution (see Table 4). Furthermore, the initial guess is very close to the calibration result.

Table 4.: calibrated SABR parameters with $\beta = 0.5$. GN denotes the Gauss-Newton minimization.

| Method	α	ρ	ν	normal vol RMSE	time(ms)
1m5y swaption in May 2014, $f = 0.0184$					
Explicit	0.052	0.404	0.837	3.19e-04	0.01
Explicit + GN	0.052	0.368	0.768	2.39e-04	0.05
Differential Evolution + GN	0.052	0.368	0.768	2.39e-04	5.03
2y5y swaption in May 2014, $f = 0.0309$					
Explicit	0.052	0.070	0.311	1.52e-05	0.01
Explicit + GN	0.052	0.058	0.313	7.59e-06	0.04
Differential Evolution + GN	0.052	0.058	0.313	7.59e-06	1.84
10y10y swaption in May 2014, $f = 0.0398$					
Explicit	0.037	-0.137	0.311	1.88e-05	0.01
Explicit + GN	0.037	-0.153	0.305	4.68e-06	0.04
Differential Evolution + GN	0.037	-0.153	0.305	4.68e-06	3.57

Figures 7, 8(a) and 8(b) show how close are the smile produced by the initial guess and the smile resulting from the Levenberg-Marquardt minimization for short expiries as well as for long ones.

6.2   Calibration on the arbitrage free PDE

We calibrate the same swaptions as in the previous section but using the arbitrage free SABR model from (Hagan *et al.*, 2014). In the calibration procedure there is the need to transform the swaption price resulting from integrating the probability density to a normal volatility, in order to compute the error in terms of normal volatilities. This can be achieved with near machine accuracy using the algorithm from (Le Floc'h, 2014). The transformation to normal volatitilies must be reasonably accurate for extremely low option prices, as even if the options reference prices are not so low, the prices resulting from the best fit procedure can be extremely low.

Short and medium expiries calibrated SABR parameters are very close to the one of the previous section (Table 5).

Figure 7.: Initial guess and calibrated smile for a May 2014 1m5y Swaption
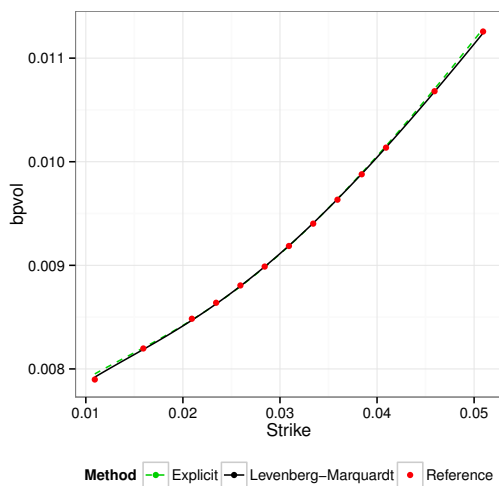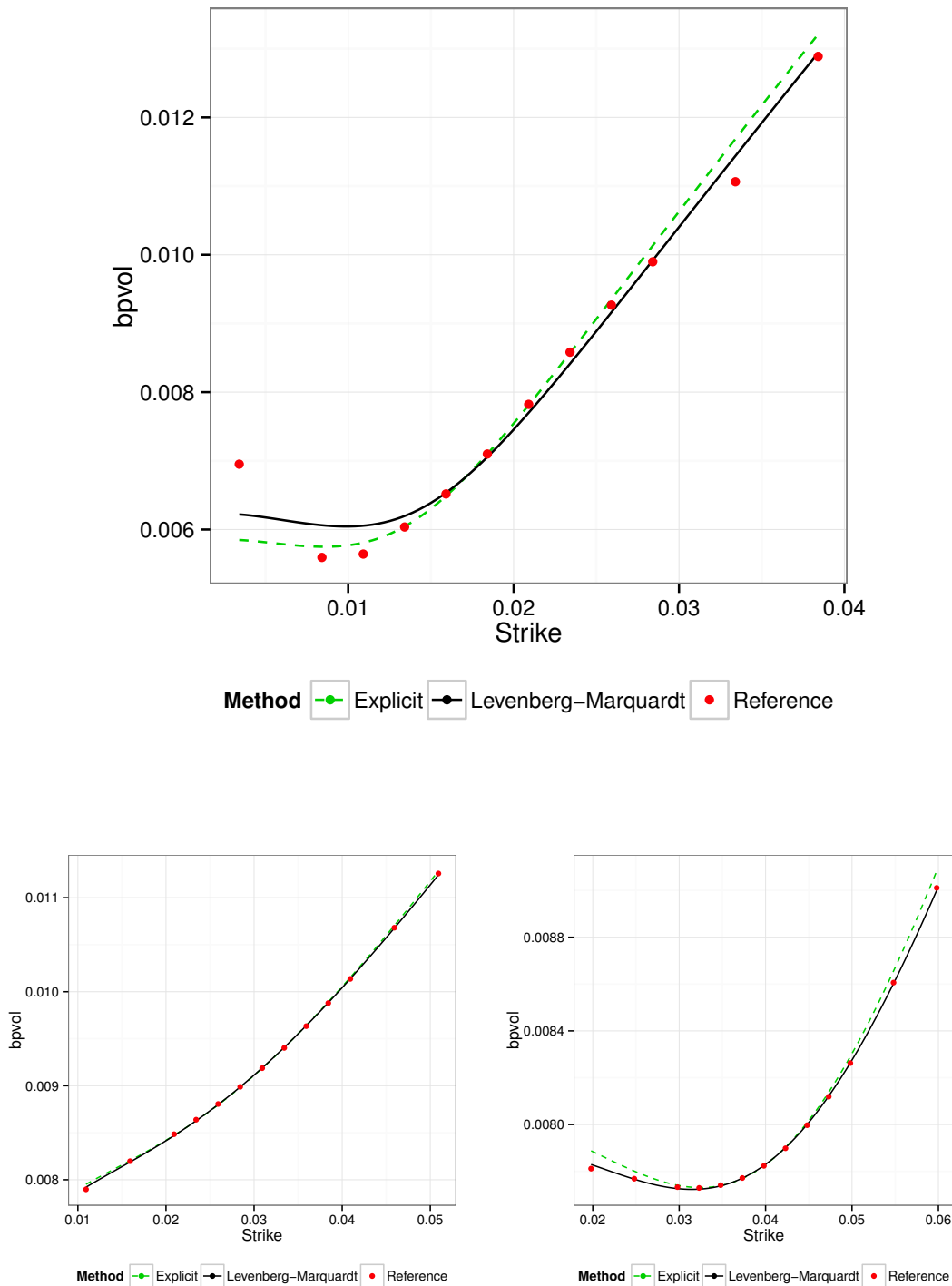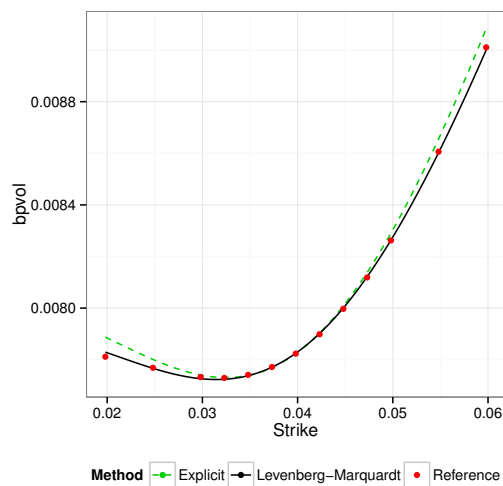
(a) 2y5y Swaption   (b) 10y10y Swaption

Figure 8.: Initial guess and calibrated smile

16   *Fabien Le Floc'h and Gary Kennedy*

There is however a visible difference for the longest expiry. In particular, the SABR parameters ρ, ν have a slightly different meaning: the smile resulting from the normal explicit initial guess is lower for low strikes with the arbitrage free PDE approach than the smile computed from the same guess with the normal formula (Figure 9).

Figure 9.: Initial guess and calibrated smile for a May 2014 10y10y Swaption using the Arbitrage-free SABR PDE approach
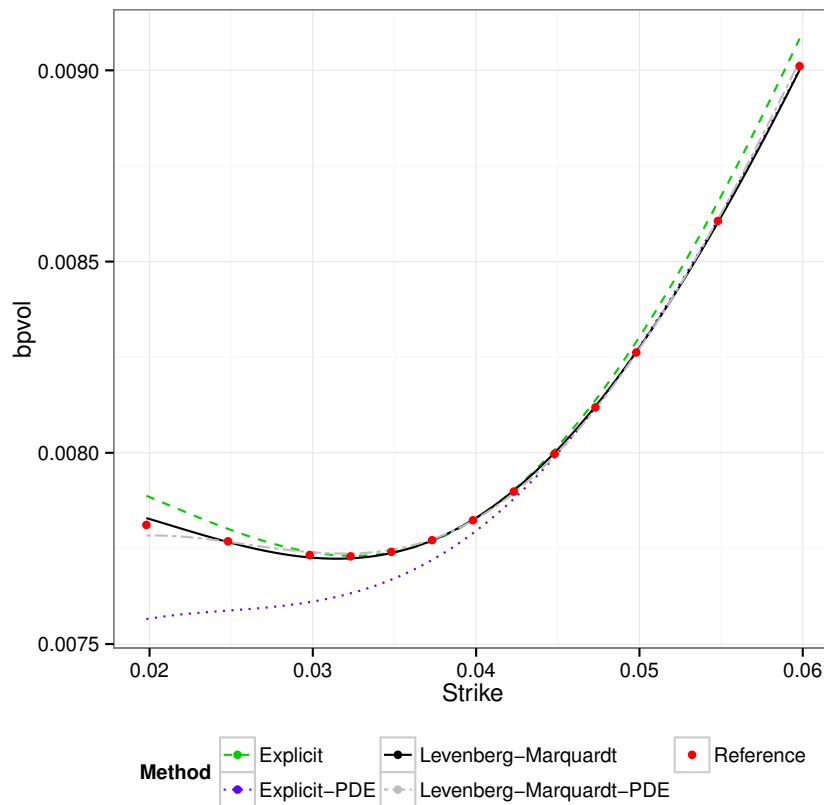

Table 5.: calibrated SABR parameters with $\beta = 0.5$ for the arbitrage free SABR model. GN denotes the Gauss-Newton minimization.

Method	α	ρ	ν	normal vol RMSE	time(s)
1m5y swaption in May 2014					
Explicit	0.052	0.404	0.837	3.13e-04	0.005
Explicit+GN	0.052	0.374	0.765	2.40e-04	0.067
2y5y swaption in May 2014					
Explicit	0.052	0.055	0.328	1.40e-05	0.005
Explicit+GN	0.052	0.058	0.323	6.61e-06	0.038
10y10y swaption in May 2014					
Explicit	0.037	-0.145	0.322	7.36e-05	0.005
Explicit+GN	0.037	-0.194	0.359	6.97e-06	0.048

7. Conclusion

We have presented an explicit initial guess procedure to calibrate the lognormal SABR formula as well as the normal SABR formula, that, despite its simplicity, works particularly well for a variety of different SABR configurations. It allows for faster calibration, by using only a local minimizer, as well as a more reliable calibration, by avoiding the problem of multiple best fits that can appear with global minimizers.

We have shown it can also be applied to the arbitrage free SABR model with success, even if the guess is further away from the target for longer maturities, due to the mismatch between the arbitrage free SABR model and the classic normal SABR formula.

The same approach to calibration, namely to recover model parameters of a second or third order expansion from a best fit parabola or cubic can be applied to other models. Such expansions can be relatively easily derived for a large class of stochastic (local) volatility models using the techniques of (Lorig *et al.*, 2014).

Appendix A. Simple expansion of the normal formula

A Taylor expansion of the normal formula (Equation 3) in $z = \log \frac{K}{f}$ and α leads to:

$$
\sigma_N(z, \alpha) = \frac{\left(\left(3 f^\beta \nu^2 \rho^2 2 f^\beta \nu^2\right) T - 24 f^\beta\right) \alpha}{24} + \frac{\beta \left(f^\beta\right)^2 \nu \rho T \alpha^2}{4 f}
$$

$$
- \frac{\left(3 f \nu^3 \rho^3 - 2 f \nu^3 \rho\right) z T - 24 f \nu \rho z}{48}
$$

$$
+ \frac{\left(\left(3 \beta f^\beta \nu^2 \rho^2 + 2 \beta f^\beta \nu^2\right) z T + 24 \beta f^\beta z\right) \alpha}{48} + \frac{\left(13 \beta^2 - 8 \beta\right) \left(f^\beta\right)^2 \nu \rho z T \alpha^2}{48 f}
$$

$$
+ \frac{\left(9 f^2 \nu^4 \rho^4 - 12 f^2 \nu^4 \rho^2 + 4 f^2 \nu^4\right) z^2 T + \left(-72 f^2 \nu^2 \rho^2 + 48 f^2 \nu^2\right) z^2}{288 f^\beta \alpha}
$$

$$
- \frac{\left((6 \beta + 3) f \nu^3 \rho^3 + (-4 \beta - 2) f \nu^3 \rho\right) z^2 T - 24 f \nu \rho z^2}{96}
$$

$$
+ \frac{\left(\left(\left(12 \beta^2 + 3 \beta\right) f^\beta \nu^2 \rho^2 + \left(4 \beta^2 - 2 \beta\right) f^\beta \nu^2\right) z^2 T + \left(24 \beta^2 + 24 \beta\right) f^\beta z^2\right) \alpha}{288}
$$

$$
+ \frac{\left(13 \beta^3 - 15 \beta^2 + 5 \beta\right) \left(f^\beta\right)^2 \nu \rho z^2 T \alpha^2}{96 f} + o(z^2) + o(\alpha^2)
$$

Assuming that ν is also small and neglecting higher order cross terms leads to our simple expansion.

Appendix B. Least squares parabola

The parabola $a + bx + cx^2$ that minimizes the square error $E = \sum_{i=1}^n w_i \left(y_i - \left(a + bx_i + cx_i^2\right)\right)^2$ at the points (x_i, y_i) with weights w_i verifies:

$$
\begin{cases}
\frac{\partial E}{\partial a} &= 2 \sum_{i=1}^n w_i \left(y_i - \left(a + bx_i + cx_i^2\right)\right) = 0 \\
\frac{\partial E}{\partial b} &= 2 \sum_{i=1}^n w_i x_i \left(y_i - \left(a + bx_i + cx_i^2\right)\right) = 0 \\
\frac{\partial E}{\partial c} &= 2 \sum_{i=1}^n w_i x_i^2 \left(y_i - \left(a + bx_i + cx_i^2\right)\right) = 0
\end{cases}
\tag{B1}
$$

The coefficients a, b, c are the solution of the following linear system:

$$
\begin{pmatrix}
\sum_{i=1}^n w_i & \sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i x_i^2 \\
\sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i x_i^2 & \sum_{i=1}^n w_i x_i^3 \\
\sum_{i=1}^n w_i x_i^2 & \sum_{i=1}^n w_i x_i^3 & \sum_{i=1}^n w_i x_i^4
\end{pmatrix}
\begin{pmatrix}
a \\
b \\
c
\end{pmatrix}
=
\begin{pmatrix}
\sum_{i=1}^n w_i y_i \\
\sum_{i=1}^n w_i x_i y_i \\
\sum_{i=1}^n w_i x_i^2 y_i
\end{pmatrix}
\tag{B2}
$$

It can be useful to pin a specific point (f, σ_f), for example to make sure that the at-the-money volatility is exactly represented by the parabola as it is available in the swaption market. In

this case the system to solve becomes:

$$\begin{pmatrix} 1 & f & f^2 \\ \sum\limits_{i=1}^{n} w_i x_i & \sum\limits_{i=1}^{n} w_i x_i^2 & \sum\limits_{i=1}^{n} w_i x_i^3 \\ \sum\limits_{i=1}^{n} w_i x_i^2 & \sum\limits_{i=1}^{n} w_i x_i^3 & \sum\limits_{i=1}^{n} w_i x_i^4 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sigma_f \\ \sum\limits_{i=1}^{n} w_i x_i y_i \\ \sum\limits_{i=1}^{n} w_i x_i^2 y_i \end{pmatrix} \tag{B3}$$

Appendix C. Moore-Penrose Gauss-Newton least squares minimization

Let x_k represent the SABR parameters (α, ρ, ν) at step k, and $r(x_k)$ the vector of the (weighted) difference between the SABR formula volatilities stemming from x_k and the market volatilities at each market strike considered in the minimization. Let $J(x_k)$ be the jacobian at x_k:

$$J(x_k) = \left(\frac{\partial r}{\partial \alpha}(x_k), \frac{\partial r}{\partial \rho}(x_k), \frac{\partial r}{\partial \nu}(x_k) \right) \tag{C1}$$

The classic Gauss-Newton method for least squares minimization consist in computing iteratively:

$$x_{k+1} = x_k - p_k \tag{C2}$$

where

$$p_k = \left(J(x_k)^T J(x_k) \right)^{-1} J(x_k)^T r(x_k)) \tag{C3}$$

It fails if the matrix $H(x_k) = J(x_k)^T J(x_k)$ is singular. In practice, because of limited machine accuracy, it fails also when H is almost singular, that is when some of its eigenvalues are close to zero.

If instead of computing an exact matrix inverse H^{-1}, we use a Moore-Penrose pseudo inverse H^+, the method will continue to work on rank-deficient problems: it will, in effect, reduce the problem to the dimensions where it is well defined. One can view this as finding the minimum on the edge: the coordinates of the x_k corresponding to near zero eigenvalues in H will stay constant.

To achieve this, we compute the Moore-Penrose inverse by singular value decomposition:

$$H^+ = V\Sigma^+ U^T \tag{C4}$$

where the singular value decomposition of H is $H = U\Sigma V^T$, and Σ^+ is computed by taking the inverse of each element in the diagonal whose absolute value is higher than a tolerance t. t is chosen to be $t = \epsilon \max(m, n) \max(\Sigma)$ where m, n are the number of rows and columns of H and ϵ is the machine epsilon, as in the Matlab/Octave function `pinv`.

For SABR calibration, the number of dimensions is just 3, and computing the Moore-Penrose pseudo inverse by singular value decomposition has no performance impact. We have found that, even when the initial guess is close to the true minimum, the system could be rank-deficient, especially for long maturity equity smiles where the optimal ρ is close to -1.

To make sure the SABR parameters don't go outside their range $\alpha \geq 0, -1 \geq \rho \geq 1, \nu \geq 0$ we rely on a smooth transformation of the domain. $\mathbb{R}$ is transformed to the interval (a, b) through the sigmoid function g defined as:

$$g(x) = a + \frac{b - a}{1 + e^{-\frac{x-a}{b-a}}} \tag{C5}$$

It would be also possible to just cap or floor the parameters (making sure the Jacobian is zero over the cap or under the floor), but while this can speed up slightly convergence, it can also have the adverse effect of being stuck too quickly on the edge.

References

Andersen, L. B. and Brotherton-Ratcliffe, R. (2005) Extended Libor Market Models with Stochastic Volatility, *Journal of Computational Finance*.

Andreasen, J. and Huge, B. (2011) ZABR–Expansions for the Masses, *Available at SSRN 1980726*.

Benaim, S., Dodgson, M. and Kainth, D., An arbitrage-free method for smile extrapolation. (2008) , Technical report, Working Paper, QuaRC, Royal Bank of Scotland.

Forde, M., Jacquier, A. and Lee, R. (2012) The small-time smile and term structure of implied volatility under the Heston model, *SIAM Journal on Financial Mathematics*, 3(1), pp. 690–708.

Gauthier, P. and Rivaille, P. Y. H. (2009) Fitting the smile, Smart parameters for SABR and Heston, *Smart Parameters for SABR and Heston (October 30, 2009)*.

Hagan, P. S. (2014) Arbitrage free SABR, *Talk kept at Global Derivatives, Amsterdam*, , pp. 14–15.

Hagan, P. S., Kumar, D., Lesniewski, A. S. and Woodward, D. E. (2002) Managing smile risk, *Wilmott magazine*.

Hagan, P. S., Kumar, D., Lesniewski, A. S. and Woodward, D. E. (2014) Arbitrage free SABR, *Wilmott magazine*.

Jäckel, P. (2013) Lets be rational. , `http://www.pjaeckel.webspace.virginmedia.com/LetsBeRational.pdf` (accessed 03-April-2014).

Johnson, S. and Nonas, B. (2009) Arbitrage-free construction of the swaption cube, *Wilmott Journal*, 1(3), pp. 137–143.

Karlsmark, M. (2013) SABR Excursions, *Four Essays in Quantitative Finance*, , p. 13.

Le Floc'h, F. (2012) Issues of Nelder-Mead Simplex Optimisation with Constraints, *SSRN eLibrary* `http://papers.ssrn.com/abstract=2097904`.

Le Floc'h, F. (2014) Fast and Accurate Analytic Basis Point Volatility, *SSRN eLibrary* `http://papers.ssrn.com/abstract=2420757`.

Le Floc'h, F. and Kennedy, G. (2014) Finite difference techniques for arbitrage free SABR, *SSRN eLibrary* `http://papers.ssrn.com/abstract=2402001`.

Levenberg, K. (1944) A method for the solution of certain problems in least squares, *Quarterly of applied mathematics*, 2, pp. 164–168.

Li, M. and Lee, K. (2011) An adaptive successive over-relaxation method for computing the Black–Scholes implied volatility, *Quantitative Finance*, 11(8), pp. 1245–1269.

Lorig, M., Pagliarani, S. and Pascucci, A. (2014) Explicit implied vols for multifactor local-stochastic vol models, *arXiv preprint arXiv:1306.5447v3*.

Marquardt, D. W. (1963) An algorithm for least-squares estimation of nonlinear parameters, *Journal of the Society for Industrial & Applied Mathematics*, 11(2), pp. 431–441.

Oblój, J. (2008) Fine-tune your smile: Correction to Hagan et al, *Wilmott Magazine*.

Paulot, L. (2009) Asymptotic implied volatility at the second order with application to the SABR model, *Available at SSRN 1413649*.

Storn, R. and Price, K. (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, *Journal of global optimization*, 11(4), pp. 341–359.

West, G. (2005) Calibration of the SABR model in illiquid markets, *Applied Mathematical Finance*, 12(4), pp. 371–385.