# Calibration of SABR Stochastic Volatility Model



November 2011

*last update: August 19, 2022*

TABLE OF CONTENTS

## 1. Introduction

SABR model is a CEV model augmented by stochastic volatility that assumes the forward rate evolves under the associated forward (terminal) measure $\mathbb{Q}^T$

$$dF_{t,T} = \varsigma_t F_{t,T}^{\beta} dW_t, \qquad d\varsigma_t = \nu \varsigma_t dZ_t, \qquad dW_t dZ_t = \rho dt \tag{1}$$

for time $t$ between initial time $s$ and maturity $T$. The $F_{t,T}$ is a forward rate process with initial value $F_{s,T} = f$. The $\varsigma_t$ is the stochastic volatility with initial value $\varsigma_s = \alpha$. The parameter $\alpha$ cannot be observed from the market, however it can be derived analytically from the at-the-money implied volatility as we shall see in due course. The factor $\nu$ is known as the volatility of volatility, which adjusts the degree of volatility clustering in time. The parameter $\beta \in [0,1]$ controls the relationship between the forward rate and the at-the-money volatility. A $\beta < 1$ ("non-lognormal" case) leads to skews in the implied volatilities. In the case of $\beta \approx 1$, if the market were to move up or down, the level of the at-the-money volatility would not be significantly affected, whereas when $\beta < 1$ the volatility increases as forward rate falls (i.e. volatility and forward move in opposite direction). The closer to 0 the more pronounced would be this effect. The correlation parameter $\rho$ plays a similar role as the $\beta$ does. It defines how the market moves in sync with the volatility dynamics. The model parameters $\nu$, $\alpha$, $\beta$ and $\rho$ are all assumed to be deterministic and time invariant.

## 2. Asymptotic Solution by Hagan et al.

Using singular perturbation techniques, Hagan et al. [1] provide a closed form asymptotic solution (up to the accuracy of a series expansion) for prices of vanilla instruments. The value of a vanilla option under the SABR model is given by the appropriate Black formula provided that the correct Black implied volatility is used. Given the forward initial value $F_{s,T} = f$ and the expiry time $\tau = T - s$, the Black implied volatility $\sigma$ can be derived as a function of strike price $K$ from a given set of SABR parameters $\alpha$, $\nu$, $\beta$ and $\rho$ by

$$\sigma_B(K,\tau;\alpha,\nu,\beta,\rho) = I_0\big(1 + I_1\tau + O(\tau^2)\big) \quad \text{where} \tag{2}$$

$$I_0 = \frac{\alpha}{p\left(1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots\right)} \cdot \frac{z}{\lambda} = \frac{\nu q}{\lambda} \cdot \frac{1}{1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots}$$

$$I_1 = \frac{(1-\beta)^2 \alpha^2}{24 p^2} + \frac{\rho \beta \nu \alpha}{4p} + \frac{(2 - 3\rho^2)\nu^2}{24}$$

$$p = (fK)^{\frac{1-\beta}{2}}, \qquad q = \log \frac{f}{K}, \qquad z = \frac{\nu p q}{\alpha}, \qquad \lambda = \log \frac{\sqrt{1 - 2\rho z + z^2} + z - \rho}{1 - \rho}$$

When $\nu \to 0$, the (2) reduces to give the at-the-money implied volatility

$$\sigma_B(K, \tau; \alpha, \nu, \beta, \rho) = I_0\big(1 + I_1 \tau + O(\tau^2)\big) \quad \text{where}$$

$$I_0 = \frac{\alpha}{p\left(1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots\right)} \cdot \frac{z}{\lambda} = \frac{\nu q}{\lambda} \cdot \frac{1}{1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots}$$

$$I_1 = \frac{(1-\beta)^2 \alpha^2}{24 p^2} \tag{3}$$

$$p = (fK)^{\frac{1-\beta}{2}}, \qquad q = \log \frac{f}{K}, \qquad z = \frac{\nu p q}{\alpha}, \qquad \lambda = \log \frac{\sqrt{1 - 2\rho z + z^2} + z - \rho}{1 - \rho}$$

When $K \to f$, the (2) reduces to give the at-the-money implied volatility

$$\sigma_f = \sigma_B(f, \tau; \alpha, \nu, \beta, \rho) = \frac{\alpha}{f^{1-\beta}}\left(1 + \left(\frac{(1-\beta)^2 \alpha^2}{24 f^{2-2\beta}} + \frac{\rho \beta \nu \alpha}{4 f^{1-\beta}} + \frac{(2 - 3\rho^2)\nu^2}{24}\right)\tau + \cdots\right) \tag{4}$$

When $\beta \to 1$, we have the implied volatility in lognormal SABR

$$\lim_{\beta \to 1} \sigma_B(K, \tau; \alpha, \nu, \beta, \rho) = I_0(1 + I_1 \tau + \cdots) \quad \text{where}$$

$$I_0 = \frac{\nu q}{\lambda}, \qquad I_1 = \frac{\rho \nu \alpha}{4} + \frac{(2 - 3\rho^2)\nu^2}{24} \tag{5}$$

using the limit $\lim_{\beta \to 1} p = 1$.

The (4) shows that there exists a relationship

$$\log \sigma_f = \log \alpha - (1 - \beta) \log f + \cdots \tag{6}$$

It indicates that the value of $\beta$ can be estimated from a log-log regression of $\sigma_f$ and $f$ with historical data by ignoring terms involving $\tau$. Alternatively, since the parameters $\beta$ and $\rho$ in SABR model control the distribution function in similar ways (i.e. both control the skewness of the distribution), the redundancy between the two parameters allows one to calibrate the model by fixing $\beta$ to an assumption (e.g. $\beta = 0.5$). The decision is often made on the basis of market experience. The remaining parameters $\alpha$, $\nu$ and $\rho$ have different effects on the volatility curve. The parameter $\alpha$ mainly controls the overall magnitude of the curve, the $\nu$ controls how much smile (i.e. convexity) the curve exhibits and the $\rho$ controls the curve's skew.

As shown in (4) the parameter $\alpha$ has a functional form with the at-the-money volatility $\sigma_f$. Inverting the equation gives the value of $\alpha$ as a root of a cubic equation if the $\nu$ and $\rho$ are known (in general, the smallest positive root would be taken if there were three real roots)

$$\frac{(1-\beta)^2\tau}{24f^{2(1-\beta)}}\alpha^3 + \frac{\rho\beta\nu\tau}{4f^{1-\beta}}\alpha^2 + \left(1 + \frac{(2-3\rho^2)\nu^2\tau}{24}\right)\alpha - \sigma_f f^{1-\beta} = 0 \tag{7}$$

This indicates that in SABR model we only need to calibrate $\rho$ and $\nu$ to implied volatility surface, providing that the value of $\beta$ is prescribed and the at-the-money implied volatility $\sigma_f$ is given. The calibration is performed at each maturity of the volatility surface by minimizing the objective function defined as a sum of squared residuals (or sum of vega weighted squared residuals)

$$\underset{\nu,\rho}{\operatorname{argmin}} \sum_{i=1}^{N} \big(\sigma_{MKT}(K_i;\tau) - \sigma_B(K_i;\tau,f,\nu,\alpha,\beta,\rho)\big)^2 \tag{8}$$

Various nonlinear optimization routines can be used to carry out the calibration, for example, Levenberg-Marquardt method or Nelder-Mead simplex method.

## 3. Obloj's Formula: Correction to Hagan et al. Solution

The general formula of the implied Black volatility for the SABR model is given by (A.65) in [1], where the general form of $\zeta$ can be found in (A.57c) in [1]. When assuming CEV model for the forward process (i.e. $C(F) = F^{\beta}$), the $\zeta$ takes the form

$$\zeta = \frac{\nu}{\alpha}\int_K^f \frac{dF}{C(F)} = \frac{\nu}{\alpha}\int_K^f \frac{dF}{F^\beta} = \frac{\nu}{\alpha}\eta, \qquad \eta = \int_K^f \frac{dF}{F^\beta} = \begin{cases} \dfrac{f^{1-\beta} - K^{1-\beta}}{1-\beta}, & \text{if } \beta < 1 \\ \log\dfrac{f}{K}, & \text{if } \beta = 1 \end{cases} \tag{9}$$

By expanding, we have

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} \Longrightarrow e^y - e^{-y} = 2y\left(1 + \frac{y^2}{3!} + \frac{y^4}{5!} + \cdots\right)$$

$$\Longrightarrow p\left(e^{\frac{1-\beta}{2}q} - e^{-\frac{1-\beta}{2}q}\right) = p(1-\beta)q\left(1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots\right),$$

$$\text{for} \quad y = \frac{1-\beta}{2}q, \qquad p = (fK)^{\frac{1-\beta}{2}}, \qquad q = \log\frac{f}{K} \tag{10}$$

$$\Longrightarrow f^{1-\beta} - K^{1-\beta} = (1-\beta)pq\left(1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots\right)$$

$$\Longrightarrow \eta = pq\left(1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots\right)$$

Thus $\zeta$ can be written as

$$\zeta = \frac{\nu}{\alpha}\eta = \frac{\nu pq}{\alpha}\left(1 + \frac{(1-\beta)^2 q^2}{24} + \frac{(1-\beta)^4 q^4}{1920} + \cdots\right) \tag{11}$$

Clearly the expression of $z$ used in (2) is just an approximation of $\zeta$ in (11) truncating all higher order terms of $q$. This leads to a correction to the original Hagan et al. solution proposed by Obloj [2] in 2008, where Obloj uses $\zeta = \frac{\nu}{\alpha}\eta$ in (9) in the general approximation formula (A.65) in [1]. This leads to the improved implied volatility formula

$$\sigma_B(K;\alpha,\nu,\beta,\rho) = I_0(1 + I_1\tau + \cdots) \quad \text{where}$$

$$I_0 = \frac{\alpha q}{\eta}\cdot\frac{\zeta}{\lambda} = \frac{\nu q}{\lambda}, \qquad I_1 = \frac{(1-\beta)^2\alpha^2}{24p^2} + \frac{\rho\beta\nu\alpha}{4p} + \frac{(2-3\rho^2)\nu^2}{24} \tag{12}$$

$$p = (fK)^{\frac{1-\beta}{2}}, \qquad q = \log\frac{f}{K}, \qquad \zeta = \frac{\nu}{\alpha}\eta, \qquad \lambda = \log\frac{\sqrt{1 - 2\rho\zeta + \zeta^2} + \zeta - \rho}{1-\rho}$$

$$\eta = \int_K^f \frac{dF}{F^\beta} = \begin{cases} \dfrac{f^{1-\beta} - K^{1-\beta}}{1-\beta}, & \text{if } \beta < 1 \\ \log \dfrac{f}{K}, & \text{if } \beta = 1 \end{cases}$$

Two asymptotic cases must be addressed. Firstly, when $K \to f$, we get the at-the-money volatility

$$\lim_{K\to f} \sigma_B(K,\tau;\alpha,\nu,\beta,\rho) = \frac{\alpha}{f^{1-\beta}}\left(1 + \left(\frac{(1-\beta)^2\alpha^2}{24 f^{2-2\beta}} + \frac{\rho\beta\nu\alpha}{4 f^{1-\beta}} + \frac{(2-3\rho^2)\nu^2}{24}\right)\tau + \cdots\right) \tag{13}$$

with the limits for $I_0$ (derived from the series expansion in (10))

$$\begin{aligned} &\lim_{K\to f}\frac{\eta}{q} = \lim_{K\to f}\frac{f^{1-\beta} - K^{1-\beta}}{(1-\beta)q} = \lim_{K\to f}\frac{-(1-\beta)K^{-\beta}}{-(1-\beta)\frac{1}{K}} = f^{1-\beta}, \qquad \lim_{K\to f}\frac{\zeta}{\lambda} = 1 \\ &\Rightarrow \lim_{K\to f} I_0 = \lim_{K\to f} \alpha \cdot \frac{q}{\eta} \cdot \frac{\zeta}{\lambda} = \frac{\alpha}{f^{1-\beta}} \end{aligned} \tag{14}$$

Secondly, when $\beta \to 1$, we have the implied volatility in lognormal SABR

$$\begin{aligned} &\lim_{\beta\to 1} \sigma_B(K,\tau;\alpha,\nu,\beta,\rho) = I_0(1 + I_1\tau + \cdots) \quad \text{where} \\ &I_0 = \frac{\nu q}{\lambda}, \qquad I_1 = \frac{\rho\nu\alpha}{4} + \frac{(2-3\rho^2)\nu^2}{24} \end{aligned} \tag{15}$$

with the limit for $\zeta$

$$\lim_{\beta\to 1}\frac{f^{1-\beta} - K^{1-\beta}}{1-\beta} = \lim_{\beta\to 1}\frac{-f^{1-\beta}\log f + K^{1-\beta}\log K}{-1} = \log\frac{f}{K} = q \Rightarrow \lim_{\beta\to 1}\zeta = \frac{\nu}{\alpha} q \tag{16}$$

## 4. Paulot's Formula: First Order Approximation

Paulot provides a first order approximation formula [3] for the implied Black volatility in the SABR model

$$\begin{aligned} &\sigma_B(K,\tau;\alpha,\nu,\beta,\rho) = I_0(1 + I_1\tau + \cdots) \quad \text{where} \\ &I_0 = \frac{\nu}{\lambda} q, \qquad I_1 = -\frac{\nu^2}{\lambda^2}\left(C + \log\frac{I_0\sqrt{fK}}{\nu}\right) \\ &q = \log\frac{f}{K}, \qquad \zeta = \frac{\nu}{\alpha}\cdot\int_K^f \frac{dF}{F^\beta} = \frac{\nu}{\alpha}\cdot\begin{cases} \dfrac{f^{1-\beta} - K^{1-\beta}}{1-\beta}, & \text{if } \beta < 1 \\ q, & \text{if } \beta = 1 \end{cases} \end{aligned} \tag{17}$$

$$\lambda = \log \frac{\sqrt{1-2\rho\zeta+\zeta^2}+\zeta-\rho}{1-\rho}$$

$$C = -\frac{1}{2}\log\frac{\alpha V f^\beta K^\beta}{\nu} + \begin{cases} -\frac{\beta\rho\big(G(t_2)-G(t_1)\big)}{(1-\beta)\sqrt{1-\rho^2}}, & \text{if } \beta < 1 \\ \frac{\rho}{2(1-\rho^2)}\Big(\frac{\alpha}{\nu} - V - \rho q\Big), & \text{if } \beta = 1 \end{cases}$$

$$V = \frac{\alpha}{\nu}\sqrt{1-2\rho\zeta+\zeta^2}$$

$$a = f^{1-\beta}, \qquad b = (1-\beta)\sqrt{1-\rho^2}, \qquad \xi = (a+bX)^2 - (1-\beta)^2 R^2$$

$$G(t) = \tan^{-1} t + \begin{cases} -\frac{a+bX}{\sqrt{\xi}}\tan^{-1}\frac{(1-\beta)\rho R + t\big(a+b(X-R)\big)}{\sqrt{\xi}}, & \text{if } \xi > 0 \\ \frac{a+bX}{(1-\beta)\rho R + t\big(a+b(X-R)\big)}, & \text{if } \xi = 0 \\ -\frac{a+bX}{\sqrt{-\xi}}\widetilde{\tanh}^{-1}\frac{(1-\beta)\rho R + t\big(a+b(X-R)\big)}{\sqrt{-\xi}}, & \text{if } \xi < 0 \end{cases}$$

$$\widetilde{\tanh}^{-1}(t) = \frac{1}{2}\log\left|\frac{1+t}{1-t}\right|$$

$$x_1 = -\frac{\rho\alpha}{\nu\sqrt{1-\rho^2}}, \qquad x_2 = -\frac{\frac{\zeta\alpha}{\nu}+\rho V}{\sqrt{1-\rho^2}}, \qquad y_1 = \frac{\alpha}{\nu}, \qquad y_2 = V$$

$$t_i = \sqrt{\frac{R-x_i+X}{R+x_i-X}}, \qquad X = \frac{x_2^2-x_1^2+y_2^2-y_1^2}{2(x_2-x_1)}, \qquad R = \sqrt{y_1^2+(x_1-X)^2}$$

Note that the $\lambda$ has a different expression than the original Paulot's formula (see Eq. 32 in [3]), however as shown below they are equivalent

$$\begin{aligned} &-\log\frac{\sqrt{\alpha^2+2\rho\alpha\nu p+\nu^2p^2}+\rho\alpha+p\nu}{(1+\rho)\alpha} \\ &= \log\frac{(1+\rho)\alpha}{\sqrt{\alpha^2+2\rho\alpha\nu p+\nu^2p^2}+\rho\alpha+p\nu}\cdot\frac{\sqrt{\alpha^2+2\rho\alpha\nu p+\nu^2p^2}-\rho\alpha-p\nu}{\sqrt{\alpha^2+2\rho\alpha\nu p+\nu^2p^2}-\rho\alpha-p\nu} \\ &= \log\frac{\sqrt{\alpha^2+2\rho\alpha\nu p+\nu^2p^2}-\rho\alpha-p\nu}{(1-\rho^2)\alpha^2}(1+\rho)\alpha = \log\frac{\sqrt{1+2\rho\frac{\nu p}{\alpha}+\frac{\nu^2p^2}{\alpha^2}}-\rho-\frac{p\nu}{\alpha}}{1-\rho} \end{aligned} \tag{18}$$

$$= \log \frac{\sqrt{1 - 2\rho\zeta + \zeta^2} + \zeta - \rho}{1 - \rho} = \lambda$$

where we define

$$p = \int_f^K \frac{dF}{F^\beta} \quad \text{and thus} \quad \zeta = \frac{\nu}{\alpha} \cdot \int_K^f \frac{dF}{F^\beta} = -\frac{\nu}{\alpha} p \tag{19}$$

For the asymptotic case when $\beta \to 1$, we have

$$\sigma_B(K, \tau; \alpha, \nu, \beta, \rho) = I_0(1 + I_1\tau + \cdots) \quad \text{where}$$

$$I_0 = \frac{\nu q}{\lambda}, \qquad I_1 = -\frac{\nu^2}{\lambda^2}\left(C + \log \frac{I_0\sqrt{fK}}{\nu}\right)$$

$$q = \log \frac{f}{K}, \qquad \zeta = \frac{\nu q}{\alpha}, \qquad \lambda = \log \frac{\sqrt{1 - 2\rho\zeta + \zeta^2} + \zeta - \rho}{1 - \rho} \tag{20}$$

$$C = -\frac{1}{2}\log \frac{fK\alpha V}{\nu} + \frac{\rho}{2(1 - \rho^2)}\left(\frac{\alpha}{\nu} - V - \rho q\right)$$

$$V = \frac{\alpha}{\nu}\sqrt{1 - 2\rho\zeta + \zeta^2}$$

When $K \to f$, it appears that the asymptotic approximation is equal to the classic formula by Hagan et al [4], that is

$$\lim_{K \to f} \sigma_B(K, \tau; \alpha, \nu, \beta, \rho) = \frac{\alpha}{f^{1-\beta}}\left(1 + \left(\frac{(1 - \beta)^2\alpha^2}{24 f^{2-2\beta}} + \frac{\rho\beta\nu\alpha}{4 f^{1-\beta}} + \frac{(2 - 3\rho^2)\nu^2}{24}\right)\tau + \cdots\right) \tag{21}$$

This is not surprising as their expansion is in fact an expansion in both maturity and moneyness (eventually of order 0 in moneyness). Although the limit for $I_1$ is not straightforward to obtain, the limit for $I_0$ is easy to derive, using the fact $\frac{\partial \zeta}{\partial K} = -\frac{\nu}{\alpha} \cdot K^{-\beta}$

$$
\begin{aligned}
\lim_{K\to f} I_0 &= \lim_{K\to f} \frac{\nu \log \dfrac{f}{K}}{\log \dfrac{\sqrt{1-2\rho\zeta+\zeta^2}+\zeta-\rho}{1-\rho}} \\
&= \lim_{K\to f} \frac{-\dfrac{\nu}{K}}{\dfrac{1-\rho}{\sqrt{1-2\rho\zeta+\zeta^2}+\zeta-\rho} \cdot \dfrac{\dfrac{-2\rho+2\zeta}{2\sqrt{1-2\rho\zeta+\zeta^2}}+1}{1-\rho} \cdot \dfrac{\partial\zeta}{\partial K}} = \lim_{K\to f} \frac{-\dfrac{\nu}{K}}{\dfrac{\partial\zeta}{\partial K}} = \frac{\alpha}{f^{1-\beta}}
\end{aligned}
\tag{22}
$$

Paulot also provides a second-order approximation. Although it is accurate in a wider region of strike prices, it requires numerical integrations for the second-order term, which defeats the purpose as an analytic approximation.

## REFERENCES

1. Hagan, P.; Kumar, D.; Lesniewski, A. and Woodward, D., “Managing smile risk”, *Wilmott Magazine*, pages 84–108, July 2002
2. Obloj, J., “Fine-tune your smile: Correction to Hagan et al.”, *Wilmott Magazine*, 35, 102–104, 2008
3. Paulot, L., “Asymptotic implied volatility at the second order with application to the SABR model”, *Large Deviations and Asymptotic Methods in Finance*, Springer (2015) 37-69, pp 54-61
4. Paulot, L., “Asymptotic implied volatility at the second order with application to the SABR model”, *Large Deviations and Asymptotic Methods in Finance*, Springer (2015) 37-69, pp 61