Reprinted from

Risk

RISK MANAGEMENT | DERIVATIVES | REGULATION

RISK.NET AUGUST 2013

SABR spreads its wings

A *Risk* Cutting Edge feature, by Alexander Antonov, Michael Konikov and Michael Spector

SABR spreads its wings

Traditional methods for the stochastic alpha beta rho model tend to focus on expansion approximations that are inaccurate in the long maturity 'wings'. However, if the Brownian motions driving the forward and its volatility are uncorrelated, option prices are analytically tractable. In the correlated case, model parameters can be mapped to a mimicking uncorrelated model for accurate option pricing. ***Alexander Antonov****,* ***Michael Konikov*** *and* ***Michael Spector*** *explain how*

The stochastic alpha beta rho (SABR) model introduced in Hagan, Lesniewski & Woodward (2001) and Hagan *et al* (2002) is widely used by practitioners to capture the volatility skew and smile effects of interest rate options. The underlying forward rate process F_t and its volatility v_t are assumed to evolve according to the system of stochastic differential equations (SDEs):

$$dF_t = F_t^{\beta} v_t dW_1 \quad (1)$$

$$dv_t = \gamma v_t dW_2 \quad (2)$$

where F_t^{β} is the local volatility function with $0 \leq \beta \leq 1$, γ is the volatility of the volatility process, and W_1 and W_2 are two standard Brownian motions under the risk-neutral measure with correlation ρ. We assume an absorbing boundary condition for F_t at zero to guarantee that it is a martingale.

The primary use of the SABR model is in volatility surface interpolation and extrapolation. Another important application is pricing constant maturity swap (CMS) products. The CMS price is calculated via integrals of European swaption prices using a static replication formula (Hagan, 2003). The integration is done over swaption strikes from zero to infinity. This means that a SABR approximation of European swaption prices must be robust and coherent for a wide range of strikes.

In the original articles (Hagan, Lesniewski & Woodward, 2001, and Hagan *et al*, 2002), the authors came up with an approximation formula for forward values of European options $\mathcal{C}(T, K) = \mathbb{E}[(F_T - K)^+]$ using Riemannian geometry and the heat kernel approach. Later on, the logic was refined by many other authors, including Berestycki, Busca & Florent (2004), Henry-Labordère (2008) and Paulot (2009).

However, the approximation quality rapidly deteriorates with time. For maturities larger than 10 years, for example, the error in implied volatility can be 1% or more even for at-the-money values. One can easily observe bad approximation behaviour for extreme strikes as well, which can prevent a valid probability density function being obtained. These undesirable properties in the distribution's tails are especially dangerous for CMS calculations by static replication.

The initial approximation formula in Hagan *et al* (2002) is used as a standard tool for volatility surface interpolation, which has led somehow to the approximation rather than the model itself becoming an industry standard. However, the model price is more coherent.

A different approach to SABR option pricing was undertaken in Islah (2009), with an exact formula in terms of a multi-dimensional integration for the zero correlation case and a conditional Bessel process approximation for non-zero correlation. Nevertheless, a practical implementation of this exact result for calibration is hardly possible – the final formula consists of a three-dimensional integration of special functions and is computationally costly.

Finally, Andreasen & Huge (2011) proposed an approximation-based one-step partial differential equation (PDE) solver. The procedure was proven to be arbitrage-free, but still only delivers a rough approximation of the theoretical SABR model.

In this article, we improve the approximations for SABR option pricing. We first give an exact formula for the zero correlation case in terms of a simple two-dimensional integration of elementary functions. The corresponding integrands have plausible asymptotics, which permit an efficient numerical implementation. Moreover, we have found a very efficient approximation in terms of one-dimensional quasi-Gaussian integration. Although an order of magnitude slower than the almost instantaneous Hagan formula, it is significantly more accurate, especially in the wings (see table A). This makes the swaption volatility cube calibration speed suitable for practical applications.

The second technical result covers a general correlation case where we propose a very accurate approximation based on a model mapping procedure. We calculate effective coefficients of a zero-correlation SABR model, the so-called mimicking model, such that its small-time asymptotics coincide with the initial non-zero correlation case. The coefficient expressions involve simple algebra without numerical integration. Then we calculate the option price using the effective zero-correlation SABR model.

Our new results provide reduced approximation error and correct behaviour in the tails of the distribution for most model parameters. When $|\rho|$ is close to one and β is close to zero, the option price can occasionally not be convex for small strikes. but this undesirable effect is much less pronounced than in previous approximations. Moreover, due to its small amplitude and localisation, it does not affect CMS pricing by static replication.

The high accuracy of our approximation is very important for dynamic SABR models, where analytic approximation used in calibration should provide results close to those used in pricing, for example, SABR Libor market models, as in Mercurio & Morini (2009) and Rebonato, McKay & White (2009).

The classical SABR model (1) and (2) uses a widely known set of five parameters $\{F_0, v_0, \beta, \gamma, \rho\}$. Usually, the initial rate F_0 is known and the other parameters $\{v_0, \beta, \gamma, \rho\}$ are subject to calibration. In some papers, v_0 is denoted as α. It is natural to impose an absorbing boundary condition on the rate behaviour at zero, which guarantees the martingale property of the rate. This also implies a non-zero probability of the rate being zero. We can transform the SABR process F_t into a stochastic volatility Bessel process Q_t via:

$$Q_t = \frac{F_t^{1-\beta}}{1-\beta} \tag{3}$$

which, with the volatility v_t, satisfies the following SDE system:

$$dQ_t = \frac{\beta}{2(\beta-1)} Q_t^{-1} v_t^2 dt + v_t dW_1 \tag{4}$$

$$dv_t = \gamma v_t dW_2 \tag{5}$$

Denote marginal probability density functions (PDFs) of the processes F_t and Q_t as $p(t, f) = \mathbb{E}[\delta(F_t - f)]$ and $p(t, q) = \mathbb{E}[\delta(Q_t - q)]$ respectively, where $\delta(x)$ denotes the Dirac delta-function. These PDFs are related by:

$$p(t,f) = p(t,q) f^{-\beta} \tag{6}$$

We consider the classical SABR model with stochastic volatility without mean-reversion. Analytical results cannot be easily adapted to mean-reverting volatility models.

Zero correlation case

Here, we assume $\rho = 0$. The SABR rate (1) is distributed as a time-changed constant elasticity of variance (CEV) process, that is:

$$F_t \sim X_{\tau_t}$$

where X_u is a CEV process with an absorbing boundary, $dX_u = X_u^\beta dW_u$, and the stochastic time τ is independent of the Brownian motion W_1 and defined as the cumulative variance:

$$\tau_t = \int_0^t v_s^2 \, ds \tag{7}$$

A distribution of the CEV process for a given time involves a modified Bessel function and can be found, for example, in Jeanblanc, Yor & Chesney (2009). The stochastic time density $p(t, \tau) = \mathbb{E}[\delta(\tau_t - \tau)]$ was found by Yor (1992) as an integral over v of the joint density $p(t, v, \tau) = \mathbb{E}[\delta(v_t - v)\delta(\tau_t - \tau)]$, which, in turn, was expressed in terms of a one-dimensional integral. The forward value of a European call option can be written as:

$$\mathcal{C}(t,K) = \mathbb{E}\left[(F_t - K)^+\right] = \mathbb{E}\left[\left(X_{\tau_t} - K\right)^+\right]$$
$$= \int_0^\infty d\tau \int_0^\infty dv p(t,v,\tau) \mathcal{C}_{cev}(\tau,K)$$

where the CEV call option forward value $\mathcal{C}_{cev}(u, K) = \mathbb{E}[(X_u - K)^+]$ can be found in Jeanblanc, Yor & Chesney (2009) and references therein. Substituting Yor's expression for the joint density $p(t, v, \tau)$, we obtain:

$$\mathcal{C}(t,K)$$
$$= 2e^{-t\gamma^2/8} \int_0^\infty \frac{dv}{v} \left(\frac{v}{v_0}\right)^{-1/2} \int_0^\infty d\tau \mathcal{C}_{cev}(\tau,K) \frac{e^{-\frac{v^2+v_0^2}{2\tau\gamma^2}}}{2\tau} \vartheta\left(\frac{vv_0}{\tau\gamma^2}, t\gamma^2\right) \tag{8}$$

where the function $\vartheta(r, t)$ is defined as:

$$\vartheta(r,t) = \frac{r}{(-2\pi i)\sqrt{2\pi t}} \int_{-\infty}^{+\infty} e^{-r\cosh\xi - \frac{(\xi+i\pi)^2}{2t}} \sinh\xi d\xi$$

A similar formula served as a basis for the approach used in Islah (2009). The CEV option values were expressed through χ^2 probability distributions, each represented by the integral of the corresponding probability density. Altogether, it included four integrations; the integration over τ was taken analytically. The final results of Islah (2009) contain triple integrals to be calculated numerically. This means the approach is computationally intensive and has convergence problems with integration over ξ, particularly at small times t, as discussed in Carr & Schroder (2004) in the context of Asian options.

But there are ways to significantly simplify expressions for option values, coming up with a double integral of elementary functions. The basic idea is to transform the option price (8) by integrating by parts with respect to τ in order to get a τ-time derivative for the CEV option price, proportional to the CEV density. Next, we use the standard contour integral to represent the underlying modified Bessel function and perform analytical integration over τ and ξ using the residue theory. We further simplify the obtained two-dimensional integral using a convenient parameterisation (see Antonov & Spector, 2012, for details).

The result is expressed via the kernel function:

$$G(t,s) = 2\sqrt{2} \frac{e^{-t/8}}{t\sqrt{2\pi t}} \int_s^\infty du \, u e^{-\frac{u^2}{2t}} \sqrt{\cosh u - \cosh s} \tag{9}$$

which is closely related to the McKean (1970) heat kernel $G_{MK}(t, s)$. Both kernels are associated with Brownian motion on the Poincare hyperbolic plane H^2. The kernel $G(t, s)$ describes the cumulative probability $P(s(x, y) > s)$ for the hyperbolic distance $s(x, y)$ on H^2 with the probability density given by the McKean kernel $G_{MK}(t, s)$, $G(t, s) = 2\pi\int_s^\infty G_{MK}(t, s') \sinh s' ds'$. G_{MK} is norm preserving, so $G(t, 0) = 1$, as can be shown directly.

The final option price formula for zero correlation is reduced to integration over the distance s:

$$\mathcal{C}(t,K) - (F_0 - K)^+$$
$$= \frac{2}{\pi}\sqrt{KF_0} \left\{ \int_{s_-}^{s_+} ds \frac{\sin(\eta\phi(s))}{\sinh s} G(t\gamma^2, s) + \sin(\eta\pi) \int_{s_+}^\infty ds \frac{e^{-\eta\psi(s)}}{\sinh s} G(t\gamma^2, s) \right\} \tag{10}$$

where:

$$\eta = \left| \frac{1}{2(\beta-1)} \right|$$

The underlying functions $\phi(s)$ and $\psi(s)$ are defined as:

$$\phi(s) = 2\arctan\sqrt{\frac{\sinh^2 s - \sinh^2 s_-}{\sinh^2 s_+ - \sinh^2 s}}$$

$$\psi(s) = 2\text{arctanh}\sqrt{\frac{\sinh^2 s - \sinh^2 s_+}{\sinh^2 s - \sinh^2 s_-}}$$

with the integration limits s_- and s_+ given by:

$$s_- = \text{arcsinh}\left(\frac{\gamma|q - q_0|}{v_0}\right) \quad \text{and} \quad s_+ = \text{arcsinh}\left(\frac{\gamma(q + q_0)}{v_0}\right)$$

Here q and q_0 are the transformed values of the spot and strike:

$$q=\frac{K^{1-\beta}}{1-\beta}\quad\text{and}\quad q_0=\frac{F_0^{1-\beta}}{1-\beta}$$

Note that the option price depends on the parameters q_0, q and v_0 through the dimensionless quantities s_- and s_+. The two-dimensional integration in formula (10) can be performed numerically in an efficient manner; the integrands are smooth functions of the parameters. Moreover, it can be shown that the function $G(t, s)$ can be closely approximated as:

$$G(t,s)\simeq\sqrt{\frac{\sinh s}{s}}e^{-\frac{s^2}{2t}-\frac{t}{8}}\left(R(t,s)+\delta R(t,s)\right) \tag{11}$$

where:

$$R(t,s)=1+\frac{3tg(s)}{8s^2}-\frac{5t^2\left(-8s^2+3g^2(s)+24g(s)\right)}{128s^4}+\frac{35t^3\left(-40s^2+3g^3(s)+24g^2(s)+120g(s)\right)}{1024s^6} \tag{12}$$

$$g(s)=s\coth s-1$$

and the correction $\delta R(t, s)$ is defined as:

$$\delta R(t,s)=e^{\frac{t}{8}}-\frac{3072+384t+24t^2+t^3}{3072} \tag{13}$$

to guarantee that $G(t, 0) = 1$. In computation, $R(t,s)$ is replaced by its fourth-order expansion for small s, as is the square root expression in (11). The effective small-time expansion (12) can be derived following Section 4.3 of Antonov & Spector (2012) and taking few other expansion terms. The technique is based on the Hagan *et al* (2002) result for the McKean kernel. Substituting the kernel G approximation (11) in the equation (10) leads to a one-dimensional integration formula. This considerably speeds up the calculation without sacrificing precision (see results below). As mentioned in the introduction, we consider $\beta \in [0, 1)$. The limiting case $\beta = 1$ requires taking careful limits in the expression (10) and will be addressed in future articles.

The general case: non-zero correlation

Heat kernel expansion. The heat kernel expansion (DeWitt, 1965) is a small-time asymptotic approximation for parabolic PDEs. This is a regular recipe for general stochastic systems to obtain PDF expansions as a fundamental solution to the Kolmogorov equation. The density $p(t, f, v)$ expansion for the SABR model was calculated in Henry-Labordère (2008) and Paulot (2009).

The marginal PDF $p(t, f)$ is obtained by integration over volatility v, which is performed with the help of the saddle-point method, implying that the main contribution is due to the 'optimal' volatility, given by:

$$v_{\min}^2=\gamma^2\delta q^2+2\rho\gamma\delta q v_0+v_0^2\quad\text{with}\quad \delta q=\frac{K^{1-\beta}-F_0^{1-\beta}}{1-\beta}$$

This gives the following small-time expansion for a call option time-value with strike K and maturity T:

$$\mathcal{C}(T,K)-(F_0-K)^+=\frac{T^{\frac{3}{2}}}{2\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{s_{\min}^2}{T\gamma^2}-\ln\frac{s_{\min}^2}{2\gamma^2}+\ln\left(K^{\beta}\sqrt{v_0v_{\min}}\right)-\mathcal{A}_{\min}\right\}\left(1+O(T)\right) \tag{14}$$

Here the optimal geodesic distance $s_{\min}$ is a function of the initial value of the rate F_0, the initial stochastic volatility value v_0 and the strike K, and is given by:

$$s_{\min}=s(q,v_{\min})=\left|\ln\frac{v_{\min}+\rho v_0+\gamma\delta q}{(1+\rho)v_0}\right|$$

The optimal parallel transport, which also depends on the strike K, is given by:

$$\mathcal{A}_{\min}=\mathcal{A}(q,v_{\min})=\frac{\beta}{2}\ln(K/F_0)+\mathcal{B}_{\min}$$

Here:

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

with:

$$\varphi_0=\arccos\left(-\frac{\delta q\gamma+v_0\rho}{v_{\min}}\right)$$

and:

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

where:

$$u_0=\frac{\delta q\gamma\rho+v_0-v_{\min}}{\delta q\gamma\sqrt{1-\rho^2}}\quad\text{and}\quad L=\frac{v_{\min}(1-\beta)}{K^{1-\beta}\gamma\sqrt{1-\rho^2}}$$

Expression (14) was derived for strikes away from the forward, that is, $|K - F_0| >> \sqrt{T}$, and a formal substitution of $K = F_0$ leads to a divergence. Paulot (2009) came up with a correct limit expansion and explained how to find the at-the-money option time-value.

One can find further details in Henry-Labordère (2008), Paulot (2009) and Antonov & Spector (2012).

Mapping to the zero-correlation SABR model. The expansion works well for small times, but for moderate and large ones it is ineffective. We use the mapping technique of Antonov & Misirpashaev (2009), which works as follows. We produce a so-called mimicking model that has the same small-time expansion for the option, and calculate the option value based on this. For example, Hagan used the Black-Scholes or normal model. Paulot has proposed the CEV process as the mimicking model. The most popular case of the Black-Scholes mimicking model $d\tilde{F}_t = \sigma\tilde{F}_t dW_t$ has the effective volatility expansion $\sigma = \sigma_0 + \sigma_1 T$, where:

$$\sigma_0=\gamma\frac{\left|\ln\frac{K}{F_0}\right|}{s_{\min}} \tag{15}$$

$$\frac{\sigma_1}{\sigma_0}=\frac{\ln\left(K^{\beta}\sqrt{v_0v_{\min}}\right)-\mathcal{A}_{\min}-\ln\sigma_0-\frac{1}{2}\ln(KF_0)}{\frac{s_{\min}^2}{\gamma^2}} \tag{16}$$

The derivation can be found in Henry-Labordère (2008) and Paulot (2009).

We will use the SABR model with zero correlation (SABR ZC) as a mimicking model. It has characteristics and asymptotics much closer to the initial SABR model than the Black-Scholes one does. The mimicking model parameters can be strike-dependent. We denote them as in SABR ZC but with a tilde. Then, using option value (14), we should match:

$$\begin{aligned}&\frac{1}{2}\frac{\tilde{s}_{\min}^2}{T\tilde{\gamma}^2}+\ln\frac{\tilde{s}_{\min}^2}{2\tilde{\gamma}^2}-\ln\left(K^{\tilde{\beta}}\sqrt{\tilde{v}_0\tilde{v}_{\min}}\right)+\tilde{\mathcal{A}}_{\min}\\&=\frac{1}{2}\frac{s_{\min}^2}{T\gamma^2}+\ln\frac{s_{\min}^2}{2\gamma^2}-\ln\left(K^{\beta}\sqrt{v_0v_{\min}}\right)+\mathcal{A}_{\min}\end{aligned} \tag{17}$$

We fix $\tilde{\gamma}$ and $\tilde{\beta}$ in the mimicking model and look for time-expansion of $\tilde{v}_0$:

$$\tilde{v}_0=\tilde{v}_0^{(0)}+T\tilde{v}_0^{(1)}+\cdots \tag{18}$$

such that the fit (17) is satisfied for both $O(T^{-1})$ and $O(T^0)$ orders. After some algebra, we get:

$$\tilde{v}_0^{(0)}=\frac{2\Phi\delta\tilde{q}\tilde{\gamma}}{\Phi^2-1} \tag{19}$$

where:

$$\Phi=\left(\frac{v_{\min}+\rho v_0+\gamma\delta q}{(1+\rho)v_0}\right)^{\frac{\tilde{\gamma}}{\gamma}}\quad\text{and}\quad\delta\tilde{q}=\frac{K^{1-\tilde{\beta}}-F_0^{1-\tilde{\beta}}}{1-\tilde{\beta}}$$

The next correction term is slightly more complicated:

$$\frac{\tilde{v}_0^{(1)}}{\tilde{v}_0^{(0)}}=\tilde{\gamma}^2\left[\frac{\frac{1}{2}\left(\beta-\tilde{\beta}\right)\ln\left(KF_0\right)+\frac{1}{2}\ln\left(v_0v_{\min}\right)}{\frac{\Phi^2-1}{\Phi^2+1}\ln\Phi}-\frac{\frac{1}{2}\ln\left(\tilde{v}_0^{(0)}\sqrt{\delta\tilde{q}^2\tilde{\gamma}^2+\tilde{v}_0^{(0)^2}}\right)-\mathcal{B}_{\min}}{\frac{\Phi^2-1}{\Phi^2+1}\ln\Phi}\right] \tag{20}$$

Let us stress that the effective volatility expansion $\tilde{v}_0=\tilde{v}_0^{(0)}+T\,\tilde{v}_0^{(1)}$, as well as the optimal quantities, volatility $v_{\min}$ and parallel transport term $\mathcal{B}_{\min}$, explicitly depend on the strike K.

Now let us come back to fixed skew $\tilde{\beta}$ and volatility-of-volatility $\tilde{\gamma}$. The approximation accuracy is quite sensitive to these parameters. A good choice based primarily on our numerical experiments is:

$$\tilde{\beta}=\beta \tag{21}$$

$$\tilde{\gamma}^2=\gamma^2-\frac{3}{2}\left\{\gamma^2\rho^2+v_0\gamma\rho(1-\beta)F_0^{\beta-1}\right\} \tag{22}$$

The intuition behind this is the following. The same power β helps with asymptotics for small strikes (see below for a detailed discussion on asymptotics). The volatility-of-volatility $\tilde{\gamma}$ choice is inspired by a fit of the at-the-money implied volatility short-time curvature, obtained as the second derivative over the at-the-money strike $K=F_0$ of the leading term of the implied volatility expansion $\sigma_0(K)$ (15). We fixed $\tilde{\gamma}$ and $\tilde{\beta}$, and calculated $\tilde{v}_0$ because the resulting option price is most sensitive to $\tilde{v}_0$, and because the main fit of the $O(T^{-1})$ terms could be explicitly solved for $\tilde{v}_0$ (19), but not for $\tilde{\gamma}$ and $\tilde{\beta}$. The at-the-money case is just the limit $K\to F_0$. The leading-order term is:

$$\left.\tilde{v}_0^{(0)}\right|_{K=F_0}=v_0$$

in this case, and the next correction is given by:

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

Its last term comes from the second derivative of the integral $\mathcal{B}_{\min}$.

Approximation procedure. Here, we summarise the approximation procedure for the call option price $\mathcal{C}(T,K)=\mathbb{E}[(F_T-K)^+]$. Given the SABR model (1) and (2) with the five parameters $\{F_0, v_0, \beta, \gamma, \rho\}$, strike K and maturity T, we come up with (strike-dependent) mimicking processes $\tilde{F}_t$ and $\tilde{v}_t$:

$$d\tilde{F}_t=\tilde{F}_t^{\tilde{\beta}}\tilde{v}_td\tilde{W}_1,\quad d\tilde{v}_t=\tilde{\gamma}\tilde{v}_td\tilde{W}_2$$

with zero correlation between the driving Brownian motions, $\mathbb{E}[d\tilde{W}_1d\tilde{W}_2]=0$. The efficient parameters are calculated as follows: the skew $\tilde{\beta}=\beta$, the volatility-of-volatility $\tilde{\gamma}$ as in (22) and the initial (strike-dependent) effective volatility $\tilde{v}_0=\tilde{v}_0^{(0)}+T\,\tilde{v}_0^{(1)}$ as in equations (19) and (20). The approximate call option price $\mathcal{C}(T,K)\simeq\mathbb{E}[(\tilde{F}_t-K)^+]$ is finally calculated by the numerical integration of expression (10).

Asymptotics

Here, we address small- and large-strike asymptotics of the marginal densities. We present the results in terms of the Bessel form of the SABR process (4) using the PDF transformation rule (6) for the initial SABR PDF. First, we start with the zero correlation case where we have the exact solution. As shown in Antonov & Spector (2012), the small q asymptotics appeared to be linear:

$$p(t,q)\sim q\quad\text{as}\quad q\to 0$$

The corresponding SABR rate density, $f=(q(1-\beta))^{1/(1-\beta)}$, behaves as follows for small values:

$$p(t,f)=p(t,q)\frac{dq}{df}\sim f^{1-2\beta}\quad\text{as}\quad f\to 0$$

A full calculation of the asymptotics for large q proved to be too involved and will be addressed elsewhere. Here, we present the leading order of the PDF (details can be found in Antonov & Spector, 2012):

$$p(t,q)\sim e^{-\frac{1}{2t\gamma^2}\ln^2\frac{2q\gamma}{v_0}}\quad\text{as}\quad q\to\infty$$

which coincides with that given by the heat kernel small-time expansion (14), where the geodesic distance $s_{\min}\sim\ln(2q\gamma/v_0)$. We doubt, however, that the pre-exponential factors of these two different limits, $q\to\infty$ and $t\to 0$, will also coincide. It is easy to see that for small β – the Gaussian case – the distribution is quite narrow even for moderate maturities. On the other hand, a log-normal case – $\beta\to 1$ – gives very fat wings.

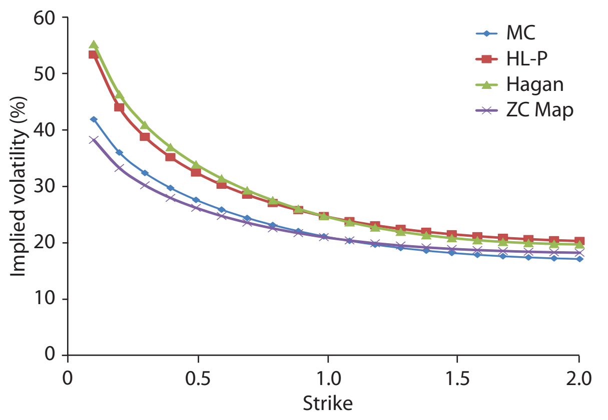
For the general correlation case, one can show that the PDF will retain linear asymptotics in q for small strikes, $p(t,q)\sim q$, and guarantee that the underlying process F_t is a global martingale. The proof is based on the forward Kolmogorov equation analysis.

For large q, the leading asymptotics corresponds to its small-time counterpart obtained by the heat kernel expansion:

$$p(t,q)\sim e^{-\frac{1}{2t\gamma^2}\ln^2\frac{2q\gamma}{(1+\rho)v_0}}\quad\text{as}\quad q\to\infty \tag{23}$$

This is an intuitive result without a strict proof, however. The cor-

1 Implied volatility error for different methods for a large maturity of 20 years



A. Implied volatility error, 20-year maturity option

Moneyness	Value (%)				Difference with MC (bp)			1D–2D (bp)
	MC	HL-P	Hagan	ZC map	HL-P	Hagan	ZC map	
10%	41.89	53.36	55.22	38.24	1,147	1,333	–365	0.3
20%	36.01	44.01	46.33	33.27	800	1032	–274	0.3
30%	32.38	38.78	40.89	30.20	640	851	–218	0.2
40%	29.72	35.18	36.97	27.96	546	725	–176	0.2
50%	27.61	32.46	33.90	26.20	485	629	–141	0.2
60%	25.88	30.29	31.40	24.76	441	552	–112	0.1
70%	24.41	28.52	29.31	23.57	411	490	–84	0.1
80%	23.16	27.04	27.54	22.57	388	438	–59	0.06
90%	22.08	25.79	26.03	21.72	371	395	–36	0.04
100%	**21.15**	**24.74**	**24.74**	**21.01**	**359**	**359**	**–14**	**–0.02**
110%	20.35	23.85	23.64	20.42	350	329	7	0.01
120%	19.67	23.10	22.72	19.92	343	305	25	0.02
130%	19.09	22.47	21.96	19.52	338	287	43	0.05
140%	18.61	21.95	21.34	19.19	334	273	58	0.04
150%	18.21	21.52	20.84	18.92	331	263	71	0.03
160%	17.88	21.17	20.46	18.71	329	258	83	0.03
170%	17.62	20.88	20.17	18.55	326	255	93	0.01
180%	17.42	20.65	19.96	18.42	323	254	100	–0.01
190%	17.25	20.47	19.81	18.32	322	256	107	0.0
200%	17.13	20.32	19.72	18.25	319	259	112	0.0

responding SABR rate probability density has the following leading asymptotics:

$$p(t,f) \sim e^{-\frac{(1-\beta)^2}{2t\gamma^2}\ln^2 f} \quad \text{as} \quad f \to \infty \tag{24}$$

This asymptotic behaviour coincides with the result of Benaim, Friz & Lee (2008) and Piterbarg (2004). The authors also derived the limit implied Black-Scholes volatility for call options $\mathcal{C}(t, K)$ for large strikes:

$$\lim_{K\to\infty} \sigma_{BS}(t,K) = \frac{\gamma}{1-\beta} \tag{25}$$

which appeared to be strike-independent. We notice also that it coincides with the large strike limit of the leading volatility term σ_0 (15).

Lee (2004) related minimum (maximum) finite moments to left (respectively, right) wings asymptotics of implied volatilities. The SABR model does not satisfy the right Lee condition: it has all positive finite moments, that is, $\mathbb{E}[F_T^p] < \infty$ for $p > 0$ and $\beta < 1$ due to its PDF strong decay (24). On the other hand, as shown in Benaim, Friz & Lee (2008), the left Lee condition is satisfied, leading to the left-wing implied volatility asymptotics:

$$\lim_{K\to 0} \frac{\sigma_{BS}^2(t,K)}{|\ln K|} = 2$$

The approximation gives a close fit for the distribution for a wide range of strikes. Nevertheless, the approximate PDF can have small negative values for small strikes, for small β and $|\rho|$ close to one. Of course, these negative values are tiny with respect to huge negative probabilities for existing approximations based on the effective implied volatility. For large strikes, our approximation appears to be close numerically to the heat kernel small-time expansion (23). We will address it rigorously elsewhere.

Numerical experiments

Here, we demonstrate the efficiency of our approach using the following data: $F_0 = 1$, $v_0 = 0.25$, $\gamma = 0.3$, $\rho = -0.5$, $\beta = 0.6$ and $T = 20$ years. We present the Black implied volatility for European call options $\mathcal{C}(T, K) = \mathbb{E}[(F_T - K)^+]$ for a range of strikes K and second-moment underlying CMS calculations.

CMS convexity adjustments depend on the second moment of the rate process, which can be evaluated by the usual static replication formula (Hagan, 2003):

$$\mathbb{E}\left[F_T^2\right] = 2\int_0^\infty dK\,\mathbb{E}\left[\left(F_T - K\right)^+\right] \tag{26}$$

For the SABR ZC map option approximation, one can use this formula directly for the second-moment calculations without any heuristic tricks, such as strike domain limitations or tail replacements. The tiny negativity of certain density approximations for the SABR ZC map does not influence the quality of the CMS calculations. For close-to-zero correlations and large skews, the big-strike tail is very fat, which produces a very slow convergence of the static replication integral.

To optimise the numerical integration, one can adapt different variance reduction techniques (for example, using the CEV model). The large strike option price can be approximated with the help of the implied volatility limit (25) with strike-independent efficient skew and volatility-of-volatility of the zero correlation SABR model.

In our numerical experiments, we compare the following methods:

- Monte Carlo simulation (MC).
- The Henry-Labordère (2008) and Paulot (2009) (HL-P) form of the Black-Scholes implied volatility expansion.
- The Hagan *et al* (2002) form of the implied volatility expansion (Hagan).
- Map to the zero-correlation SABR model (ZC map).

For the Monte Carlo simulations, we have used 100 time-steps a year, 50,000 paths of good low-discrepancy numbers and an Euler scheme with an absorbing condition for a zero rate. The simulation results are presented as the mean over 50 independent runs after a careful convergence study in both time-steps and paths.

The computer time for the ZC map approximation is almost entirely spent in the numerical two-dimensional integration (9) and (10). Using efficient high-order integration schemes allows us to obtain the ZC map approximation 100 times slower than the almost instantaneous classical Hagan one. However, one-dimensional integration with the kernel $G(t, s)$ approximation (10) and (11) slows calculation by a factor of 10 compared with the Hagan formula, due to quasi-Gaussian nature of the former. This makes the swaption volatility cube calibration speed acceptable for practical applications. Note that the error between two-dimensional integration and one-dimensional one is tiny, at most 0.3 basis points in the implied volatility. We present it in our numerical experiments under the heading 1D–2D.

When the new formula's slowness presents a bottleneck, one can use hardware to accelerate, for example, graphics processing units. Another way to speed up calibration is to find a solution with the original Hagan formula where it is known to be accurate – for example, for tiny maturities and close to at-the-money strikes – or use it as an initial guess for final calibration with a more precise new formula.

In figure 1 and table A, we present the implied volatility and its error for different methods for a large maturity of 20 years.

We observe an excellent approximation quality around the at-the-money region for the SABR ZC map, with only slight degeneration on the edges and insufficient approximation accuracy for the other methods.

Table B demonstrates an excellent approximation quality for the SABR ZC map and insufficient accuracy for the other methods. This means that our approximation works correctly even for extreme strikes. Indeed, for a 20-year maturity the second-moment integration (26) goes quite far in strikes: the option price reaches 10^{-6} for strikes around 35.

B. Centred second-moment $\mathbb{E}[(F_T - F_0)^2]$ and its errors for different methods

Value				Difference with MC			1D–2D
MC	HL-P	Hagan	ZC map	HL-P	Hagan	ZC map	
1.028	1.255	1.733	1.065	0.227	0.705	0.037	–0.001

Conclusion

The commonly used Hagan expansion for the SABR model is well known to be imprecise in the distribution's tails, and in pricing longer expiry options, implying negative densities and arbitrage. Most known alternatives either exhibit similar behaviour, not consistent with the theoretical SABR model, or have an extremely slow numerical implementation. The approach presented here is quite precise and near arbitrage-free for all practical purposes, consistent with the theoretical SABR, and still reasonably fast. There is a new exact option pricing formula for the zero-correlation case, and the general case is handled by mapping the model parameters into an uncorrelated version without much loss of precision. Although there is a reduction in computation speed of an order of magnitude, the accuracy gained is significant. ■

Alexander Antonov is senior vice-president of quantitative research at Numerix in Paris. Michael Konikov is executive director of quantitative development and Michael Spector is director of quantitative research at Numerix in New York. They are indebted to Serguei Mechkov for discussions and numerical implementation help as well as to their colleagues at Numerix, especially Gregory Whitten and Serguei Issakov for supporting this work and Patti Harris and Nic Trainor for the thorough editing. They are also grateful to Alexander Lipton for stimulating discussions. Email: antonov@numerix.com, mkonikov@numerix.com and mspector@numerix.com

References

Andreasen J and B Huge, 2011
ZABR – expansions for the masses
SSRN paper

Antonov A and T Misirpashaev, 2009
Projection on a quadratic model by asymptotic expansion with an application to LMM swaption
SSRN paper

Antonov A and M Spector, 2012
Advanced analytics for the SABR model
SSRN paper

Benaim S, P Friz and R Lee, 2008
The Black Scholes implied volatility at extreme strikes
In Frontiers in Quantitative Finance: Volatility and Credit Risk Modeling, Wiley

Berestycki H, J Busca and I Florent, 2004
Computing the implied volatility in stochastic volatility models
Communications on Pure and Applied Mathematics 57(10), pages 1,352–1,373

Carr P and M Schroder, 2004
Bessel processes, the integral of geometric Brownian motion, and Asian options
Theory of Probability and its Applications 48, pages 400–425

DeWitt B, 1965
Dynamical theory of groups and fields
Gordon and Breach

Hagan P, 2003
Convexity conundrums: pricing CMS swaps, caps, and floors
Wilmott Magazine, March, pages 38–44

Hagan P, D Kumar, A Lesniewski and D Woodward, 2002
Managing smile risk
Wilmott Magazine 3, pages 84–108

Hagan P, A Lesniewski and D Woodward, 2001
Probability distribution in the SABR model of stochastic volatility
Working paper

Henry-Labordère P, 2008
Analysis, geometry, and modeling in finance: advanced methods in option pricing
Chapman & Hall

Islah O, 2009
Solving SABR in exact form and unifying it with Libor market model
SSRN paper

Jeanblanc M, M Yor and M Chesney, 2009
Mathematics methods for financial markets
Springer

Lee R, 2004
The moment formula for implied volatility at extreme strikes
Mathematical Finance 14(3), pages 469–480

McKean H, 1970
An upper bound to the spectrum of Δ on a manifold of negative curvature
Journal of Differential Geometry 4, pages 359–366

Mercurio F and M Morini, 2009
Joining the SABR and Libor models together
Risk March, pages 80–85, available at www.risk.net/1500241

Paulot L, 2009
Asymptotic implied volatility at the second order with application to the SABR model
SSRN paper

Piterbarg V, 2004
Implied volatility smile asymptotics when all moments are finite
Working paper

Rebonato R, K McKay and R White, 2009
The SABR/Libor market model: pricing, calibration and hedging for complex interest-rate derivatives
John Wiley & Sons

Yor M, 1992
On some exponential functionals of Brownian motion
Advances in Applied Probability 24(3), pages 509–531