Asymptotic Implied Volatility at the Second Order
With Application to the SABR Model

Louis Paulot

Sophis Technology
24–26 place de la Madeleine, 75008 Paris, France

`louis.paulot@sophis.net`

June 2009

Abstract

We provide a general method to compute a Taylor expansion in time of implied volatility for stochastic volatility models, using a heat kernel expansion. Beyond the order 0 implied volatility which is already known, we compute the first order correction exactly at all strikes from the scalar coefficient of the heat kernel expansion. Furthermore, the first correction in the heat kernel expansion gives the second order correction for implied volatility, which we also give exactly at all strikes. As an application, we compute this asymptotic expansion at order 2 for the SABR model.

Keywords: Stochastic volatility, Asymptotic expansion, Heat kernel, SABR.

Contents

1 Introduction

The most known model for pricing derivatives is the Black-Scholes-Merton model, where the underlying is supposed to follow a geometric Brownian motion. Popular extensions include local volatility models and stochastic volatility models. For example the SABR model [HKLW02] combines the local volatility of the CEV model [Cox75] and a lognormal volatility process. Closed formulas for European options can be obtained for a few models; it is the case of the CEV model or for a stochastic volatility example the Heston model [Hes93]. These are however special cases and there are generally no closed form formulas. Finite difference methods or Monte-Carlo simulations can be used to price derivatives. Approximations has also been computed to achieve faster pricing, especially for calibration processes.

For short maturities, Hagan, Kumar, Lesniewski and Woodward provide an approximation for the implied volatility of the SABR model they introduce [HKLW02]. Berestycki, Busca and Florent [BBF02, BBF04] and Henry-Labordère [HL05] give general methods to compute short maturity asymptotics of stochastic volatility models. These expansions give the implied volatility at first order in maturity. In addition some quantities are approximated by their value at the money, which can produce errors in the wings of the distributions.

In this paper, we leverage on the heat kernel methods introduced for the study of stochastic processes by Varadhan [Var67a, Var67b] and used for the SABR model in [HLW05] and [HL05]. Using the heat kernel expansion of DeWitt [DeW65], we provide a method to compute exactly a Taylor

expansion of the implied volatility at all strikes. The stochastic volatility diffusion is formulated as a diffusion on a Riemannian manifold. The geodesic distance gives the implied volatility at null maturity. The multiplicative factor of the heat kernel expansion provides the first order (in time) correction to implied volatility. The first corrective term of the heat kerned is translated into the second order correction to implied volatility and similarly for higher order corrections. We perform a detailed computation up to order 2 of the Taylor expansion in time of implied volatility, without other approximations.

More generally, our method can be used to approximate a stochastic volatility model by an other model for which a closed form solution exists, with an implied parameter computed as a Taylor expansion.

As an application, we compute the asymptotic SABR volatility at order 2 and compare it to finite difference method results and to the original SABR expansion.

Our results can be useful for pricing short maturities options or even long maturities options with very low volatility of volatility. When the approximation is not valid, a numerical method such as a finite difference method (FDM) has to be used. When it is valid, it gives much faster results. At very short maturities, the prices are even more precise. Calibration at short maturities appears to be more stable using this approximation.

In section 2 we recast the financial model in physical and geometric terms and fix our conventions. In section 3 we use a heat kernel expansion to compute a short maturity expansion of Black or more generally CEV implied volatility. Finally in section 4 we apply the method to the SABR model and compare the results to FDM and to the original formula.

2 Diffusion equation in covariant form

A stochastic volatility model for some asset with pure diffusion (no jumps) is described by two risk-neutral processes: the asset price S and a variable V which describes the stochastic part of volatility. In Heston model V would be the variance whereas in the SABR model it is a factor of volatility. The diffusion is given by the stochastic differential equations

$$
\begin{aligned}
\mathrm{d}S &= \mu_S(S)\mathrm{d}t + \sigma_S(S,V)\mathrm{d}W_1 \\
\mathrm{d}V &= \mu_V(V)\mathrm{d}t + \sigma_V(V)\mathrm{d}W_2
\end{aligned}
\tag{1}
$$

where $\mathrm{d}W_1$ and $\mathrm{d}W_2$ are two standard Brownian processes with correlation ρ. The dependence of parameters in variables S and V we have written is the more common, it may be more general with all parameters depending on both variables.

Stochastic volatility models can be seen as diffusion on a Riemann surface. More precisely, prices of securities are sections of a line bundle over this Riemann surface which are solutions of a diffusion (or heat) equation.

A introduction to this subject and its applications to finance can be found in [HL08]. We present here the formalism and define all quantities we use in order to set our conventions.

2.1 Diffusion equation

Let us consider a general model with n state variables $X^i(t)$ (which will be the spot and the volatility) which follow a pure diffusion process, without jumps. For simplicity we consider a European payoff of some maturity T. The price $F(X(t),t)$ of such a payoff is the solution of a diffusion equation

$$
-\partial_t F = \mu^i \partial_i F + \frac{1}{2}\Sigma^{ij}\partial_i\partial_j F - rF
\tag{2}
$$

where Σ^{ij} is the covariance matrix, μ_i the drifts and r the numéraire rate. All coefficients can depend on state variables $X^i(t)$ and time t. Unless explicitly staten, we take Einstein sum convention: repeated indices are summed. The price of the European option is given by the solution of this equation with terminal boundary condition at maturity T given by the payoff.

The covariance matrix Σ^{ij} can be seen mathematically as the inverse $g^{ij} = \Sigma^{ij}$ of a metric g_{ij} on the space of variables. The diffusion equation describes the diffusion over a Riemannian manifold: the state of variables endowed with the metric $g_{ij} = (\Sigma^{-1})_{ij}$.

Examples

1. The Black-Scholes equation in the monetary account numéraire with volatility σ, risk free rate r and dividend yield q reads

$$-\partial_t F = (r - q)S\partial_S F + \frac{1}{2}\sigma^2 S^2 \partial_S \partial_S F - rF$$

   This is equation (2) with $\mu^S = (r - q)S$, $\Sigma^{SS} = \sigma^2 S^2$ and $r = r$.

2. For the stochastic volatility model described by equations (1), μ^i is a two-dimensional vector

$$\mu = \begin{pmatrix} \mu_S \\ \mu_V \end{pmatrix} .$$

   The covariance matrix Σ^{ij} is

$$\Sigma = \begin{pmatrix} \sigma_S{}^2 & \rho\sigma_S\sigma_V \\ \rho\sigma_S\sigma_V & \sigma_V{}^2 \end{pmatrix} .$$

In what follows we restrict ourselves to the case of time-homogeneous models: there is no explicit time-dependence in parameters. The generalization to time-dependent cases is not difficult.

2.2 Gauge structure

There are several *gauge* transformations which are natural for such systems:

1. Change of numéraire:
$$F(X,t) \longrightarrow \frac{F(X,t)}{\Phi(X,t)}$$

   where $\Phi(X,t)$ is the price of a security which is always nonzero. Mathematically it is a real function which is positive everywhere, that we denote thus by $\Phi(X,t) = e^{-\phi(X,t)}$.

2. Change of variables
$$X \longrightarrow X'(X) .$$

The natural way to handle a system with gauge freedom is to introduce covariant derivatives. The coordinate freedom is handled through the *Levi-Civita* connection which acts respectively on scalars, vectors and 1-forms as

$$\begin{aligned} D_i f &= \partial_i f \\ D_i f^j &= \partial_i f^j + \Gamma^j_{ik} f^k \\ D_i f_j &= \partial_i f_j - \Gamma^k_{ij} f_k \end{aligned}$$

4

where Γ^j_{ik} are the *Christoffel symbols*. The action on tensors with more indices is obtained by acting on all indices with the Christoffel symbols. Christoffel symbols can be computed from the metric as

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl} \left(\partial_i g_{lj} + \partial_j g_{il} - \partial_l g_{ij} \right) \ .$$

A fundamental property of the Levi-Civita connection is the covariance of the metric:

$$D_i g_{jk} = 0 \ .$$

The metric is used to transforms vectors into 1-forms and conversely, *i.e.* lowering or raising indices:

$$\begin{aligned} A^i &= g^{ij} A_j \\ A_i &= g_{ij} A^j \ . \end{aligned}$$

The numéraire gauge freedom is handled through a $\mathbb{R}$-valued connection[1] with spatial and time components given by a 1-form A_i and a scalar[2] Q:

$$\begin{aligned} \nabla_i F &= (D_i - A_i) F \\ \nabla_t F &= (\partial_t - Q) F \ . \end{aligned}$$

Under the change of numéraire

$$F \ \longrightarrow \ e^{\phi(X,t)} F$$

these operators are covariant,

$$\begin{aligned} \nabla_i F &\longrightarrow \ e^{\phi(X,t)} \nabla_i F \\ \nabla_t F &\longrightarrow \ e^{\phi(X,t)} \nabla_t F \ , \end{aligned}$$

provided that A_i and Q are shifted as

$$\begin{aligned} A_i &\longrightarrow \ A_i - \partial_i \phi \\ Q &\longrightarrow \ Q - \partial_t \phi \ . \end{aligned}$$

Using these connections, the diffusion equation (2) can be rewritten as

$$-\nabla_t F = \frac{1}{2} \nabla^i \nabla_i F \ . \tag{3}$$

Identifying terms between equations (2) and (3), the $\mathbb{R}$ connection must be

$$A_i = g_{ij} \left(-\frac{1}{2} \Gamma^j_{kl} g^{kl} - \mu^j \right) \tag{4}$$

$$Q = -\frac{1}{2} g^{ij} \left(\partial_i A_j - A_i A_j - \Gamma^k_{ij} A_k \right) + r \ . \tag{5}$$

In addition with

$$g^{ij} = \Sigma^{ij}$$

this translates the set of financial parameters into mathematical quantities.

---

[1] This connection is similar to the connection which described the electromagnetic potential, except that the fibre of the gauge bundle is $\mathbb{R}$ instead of $U(1)$. This causes a difference of a factor i in equations.

[2] There is a breaking of symmetry between time and spatial directions. The diffusion equation can be seen as a non-relativistic limit of a pure wave equation.

2.3 Kolmogorov forward equation

The Kolmogorov backward equation (3) leads to a dual Kolmogorov forward equation.

We suppose that all prices are expressed with respect to a numéraire which is a traded asset that does not pay any coupon or dividend. The price of the numéraire security itself is identically 1; this reads mathematically

$$-\nabla_t 1 = \frac{1}{2}\nabla^i\nabla_i 1$$

If $p(X, t)$ is a risk-neutral probability density to get in state X at time t starting from state X_0 at time 0, then the price of a European payoff of maturity $T \geq t$ can be written as

$$F(X_0, 0) = \int \mathrm{d}X p(X, t) F(X, t)$$

As t does not appear on the left-hand side, the derivative of the integral with respect to t must vanish.

$$\int \mathrm{d}X \partial_t(p(X, t) F(X, t)) = 0 \ .$$

We define an action of the gauge group on p with a plus sign instead of a minus sign when acting on F:

$$
\begin{aligned}
\nabla_i p &= (D_i + A_i)p \\
\nabla_t p &= (\partial_t + Q)p \ .
\end{aligned}
$$

This means that they p and F have opposite charges under the $\mathbb{R}$ gauge group, such that pF is neutral and $\nabla_t(pQ) = \partial_t(pQ)$. We have thus

$$\int \mathrm{d}X \left(\nabla_t p(X, t) F(X, t) + p(X, t)\nabla_t F(X, t)\right) = 0$$

Using equation (3) for $\nabla_t F(X, t)$ and integrating by part on the spatial directions, this equation becomes

$$\int \mathrm{d}X \left(\nabla_t p(X, t) - \frac{1}{2}\nabla^i\nabla_i p(X, t)\right) F(X, t) = 0 \tag{6}$$

This equation will be automatically satisfied if

$$\nabla_t p = \frac{1}{2}\nabla^i\nabla_i p \ . \tag{7}$$

Moreover, if the market is complete equation (6) must be true for all functions $F(\cdot, t)$ which imposes equation (7). This is the Kolmogorov forward equation, written in a covariant way. It should be noted that $p(X, t)$ is a density, which means that the Levi-Civita connection does not reduce to a partial derivative as would be the case for a scalar.

3 Asymptotic implied volatility

We consider a stochastic volatility model where the variable is a forward price or rate F with a volatility variable V:

$$
\begin{aligned}
\mathrm{d}F &= \sigma_F(F, V)\mathrm{d}W_1 \\
\mathrm{d}V &= \mu_V(V)\mathrm{d}t + \sigma_V(V)\mathrm{d}W_2
\end{aligned}
\tag{8}
$$

with $\langle \mathrm{d}W_1 \mathrm{d}W_2 \rangle = \rho \mathrm{d}t$.

Our computation of an asymptotic expansion at short time of implied volatility at strike K involves three steps:

1. Compute an asymptotic value of the transition probability from initial state F_0, V_0 at time 0 to K, V at time t using a heat kernel expansion;

2. Compute $\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right]$ using a saddle point method;

3. Integrate over time to compute the time value;

4. Compare to the same formula for the Black-Scholes model to extract the implied volatility.

3.1   Heat kernel expansion

At short time the solution of equation (7) with initial condition $p(X,0) = \delta(X-X_0)$ is asymptotically given by the heat kernel expansion [DeW65] (see [Vas03] for a review)[3]

$$p(X,t) = \frac{\sqrt{g(X)}}{(2\pi t)^{n/2}} \sqrt{\Delta(X_0, X)} \mathcal{P}(X_0, X) e^{-\frac{d^2(X_0, X)}{2t}} \sum_{k \geq 0} a_k(X_0, X) t^k \tag{9}$$

$g(X)$ is the determinant of the metric at point X:

$$g = \mathrm{Det}(g_{ij}) \ .$$

$d(X_0, X)$ is the geodesic distance between the starting point X_0 and the end point X, this is the minimal distance between X_0 and X. It can also be written as

$$\frac{d^2(X_0, X)}{t} = \min_{X(s)} \int_0^t \mathrm{d}t \, g_{ij} \dot{X}^i \dot{X}^j$$

where the minimum is taken on all paths going from $X(0) = X_0$ to $X(t) = X$. (This is independent of t.) We denote by $\mathcal{C}$ this geodesic path. $\Delta(X_0, X)$ is the Van Vleck–Morette determinant

$$\Delta(X_0, X) = \frac{\mathrm{Det}\left(-\frac{1}{2}\frac{\partial^2 d^2(X_0, X)}{\partial X^i \partial X_0^j}\right)}{\sqrt{g(X_0)g(X)}}$$

$\mathcal{P}(X_0, X)$ is the parallel transport along the geodesic with respect to the $\mathbb{R}$ connection. It is such that its covariant derivative along the geodesic path is null:

$$\mathcal{P}(X_0, X) = e^{-\int_{\mathcal{C}} A_i \dot{X}^i \mathrm{d}t} = e^{-\int_{\mathcal{C}} A_i \mathrm{d}X^i}$$

---

[3] Using Feynman path integral, the solution to equation (7) can be written up to some normalization factor as

$$p(X,t) \propto \int [DX] e^{-\int_0^t \mathrm{d}t \left(\frac{1}{2}g_{ij}\dot{X}^i \dot{X}^j + A_i \dot{X}^i + Q\right)}$$

where $[DX]$ means integrating over all path $X(s)$ going from $X(0) = X_0$ to $X(t) = X$. The normalization factor is the inverse of the same quantity with the integral computed over all paths with starting point X_0, so that the total probability is 1. It is generally not possible to compute this integral exactly. However it gives some hints on the asymptotic solution at short time: the solution will be dominated by the path corresponding to the minimal value of the integrand inside the exponential, which will be close to the geodesic path.

where the integral is computed on the geodesic path $\mathcal{C}$. Finally, $a_i(X_0, X)$ are functions which are defined recursively with

$$a_0 = 1$$

and a_i's satisfy the differential equations

$$(k + (\nabla^i d) d \nabla_i) a_k = \mathcal{P}^{-1} \Delta^{-1/2} \left(\frac{1}{2} \nabla^i \nabla_i - Q \right) \mathcal{P} \Delta^{1/2} a_{k-1} .$$

Along a given geodesic curve parameterized by its geodesic distance from the origin d, this equation reads

$$(k + d \partial_d) a_k = \mathcal{P}^{-1} \Delta^{-1/2} \left(\frac{1}{2} \nabla^i \nabla_i - Q \right) \mathcal{P} \Delta^{1/2} a_{k-1}$$

which can be integrated as

$$a_k = \frac{1}{d^k} \int_0^d ds\, s^{k-1} \mathcal{P}^{-1} \Delta^{-1/2} \left(\frac{1}{2} \nabla^i \nabla_i - Q \right) \mathcal{P} \Delta^{1/2} a_{k-1} .$$

Note that when acting with the covariant derivative, the full charge under the $\mathbb{R}$ gauge group is carried by the parallel transport term $\mathcal{P}$ and that $\frac{p(X,t)}{\sqrt{g(X)}}$ is a scalar with respect to the Levi-Civita connection. a_k's are scalars with respect to both connections.

In order to produce a first order expansion of the implied volatility, only the common multiplicative factor of expansion 9 is needed. In order to compute a second order term for the implied volatility, we will also make use of the first corrective term $a_1 t$ with

$$a_1 = \frac{1}{d} \int_0^d ds\, \mathcal{P}^{-1} \Delta^{-1/2} \left(\frac{1}{2} \nabla^i \nabla_i - Q \right) \mathcal{P} \Delta^{1/2} . \tag{10}$$

3.2 Expected variance

We now compute $\mathbb{E}\left[\sigma_F^2(K, V) \delta(F - K) \right]$. This quantity can be written as an integral over the terminal volatility variable V:

$$\mathbb{E}\left[\sigma_F^2(K, V) \delta(F - K) \right] = \int dV\, \sigma_F^2(K, V) p(K, V; t)$$

where $p(F, V; t)$ is given by the heat kernel expansion (9) with $X = \begin{pmatrix} F \\ V \end{pmatrix}$ and $n = 2$. The integrand can written as

$$\sigma_F^2(K, V) p(K, V; t) = \frac{1}{2\pi t} e^{-\frac{B}{t} - C - Dt} + o(t) \tag{11}$$

with

$$B = \frac{1}{2} d(F_0, V_0; K, V)^2 \tag{12}$$

$$C = -2 \ln(\sigma_F(K, V)) - \frac{1}{2} \left[\ln(g(K, V)) + \ln(\Delta(F_0, V_0; K, V)) \right] + \mathcal{A}(K, V) \tag{13}$$

$$D = -a_1(K, V) \tag{14}$$

where $\mathcal{C}$ is the geodesic curve joining (F_0, V_0) to (K, V), $\mathcal{A}$ is the integral of the $\mathbb{R}$ connection

$$\mathcal{A}(K, V) = -\ln(\mathcal{P}(K, V)) = \int_{\mathcal{C}} A_i dX^i$$

and a_1 is given above as an integral over the geodesic path $\mathcal{C}$ in equation (10).

The integral over (11) will be dominated at short time by the B term. More precisely, it will be dominated by the volatility $V_{\min}$ which minimizes $B(K,V) = \frac{1}{2}d(F_0, V_0; K, V)^2$. This is the final volatility which minimizes the distance between the initial conditions and the strike K. Expanding all functions in the neighborhood of $V_{\min}$, where $B'(V_{\min}) = 0$, the integrand can expanded as

$$\sigma_F^2(K,V)p(K,V;t) = \frac{1}{2\pi t}e^{-\frac{B}{t} - C - Dt}e^{-\frac{1}{2}B''\delta V^2}$$
$$\left[1 - \frac{1}{2}\left(C'' - C'^2\right)\delta V^2 - \left(\frac{1}{24}B^{(4)} - \frac{1}{6}B^{(3)}C'\right)\frac{\delta V^4}{t} + \frac{1}{72}B^{(3)2}\frac{\delta V^6}{t^2} + o(t) + \text{odd terms}\right]$$

where derivatives are with respect to V, where all functions B, C, D an their derivatives are taken at $(K, V_{\min})$ and $\delta V = V - V_{\min}$. When writing $o(t)$, we have anticipated that after integration $\delta V^2 \sim \frac{1}{t}$. We have also anticipated that odd terms in δV will not give contributions to the integral.

Integrating over δV, and using that the first even moments of the standard normal distribution are $M_2 = 1$, $M_4 = 3$ and $M_6 = 15$, we get for the integral

$$\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right] = \frac{1}{\sqrt{2\pi t B''}}e^{-\frac{B}{t} - C - Dt}$$
$$\left[1 - \frac{1}{2}\left(C'' - C'^2\right)\frac{t}{B''} - \left(\frac{1}{24}B^{(4)} - \frac{1}{6}B^{(3)}C'\right)\frac{3t}{B''^2} + \frac{1}{72}B^{(3)2}\frac{15t}{B''^3} + o(t)\right]$$

This can be rewritten as

$$\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right] = \frac{1}{\sqrt{2\pi t}}e^{-\frac{B}{t} - \widetilde{C} - \widetilde{D}t + o(t)} \tag{15}$$

with

$$\widetilde{C} = C + \frac{1}{2}\ln(B'') \tag{16}$$

$$\widetilde{D} = D + \frac{1}{2B''}\left[C'' - C'^2 + \frac{1}{4}\frac{B^{(4)}}{B''} - \frac{B^{(3)}}{B''}C' - \frac{5}{12}\left(\frac{B^{(3)}}{B''}\right)^2\right]$$

$$= D + \frac{1}{2B''}\left[\widetilde{C}'' - \widetilde{C}'^2 - \frac{1}{4}\frac{B^{(4)}}{B''} + \frac{1}{3}\left(\frac{B^{(3)}}{B''}\right)^2\right] \tag{17}$$

where all derivatives are with respect to V and all functions and their derivatives are taken at $(K, V_{\min})$.

3.3 Time value

The price of a Call of maturity T and strike K can be written as the payoff integrated against the risk-neutral distribution:
$$\text{Call}(K,T) = e^{-rT}\int dF\,(F-K)^+ p(F,T)\,.$$

This can be written also as a double integral over strike and time as

$$\text{Call}(K,T) = e^{-rT}\left[(F_0 - K)^+ + \int dF \int_0^T dt\,(F-K)^+ \partial_t p(F,T)\right]\,. \tag{18}$$

As F is a forward, driftless, the Kolmogorov forward equation reduces to

$$\partial_t p(F,t) = \frac{1}{2}\partial_F^2\left(\sigma_F^2(F,t)p(F,t)\right) \tag{19}$$

where $\sigma_F^2(F,t)$ is the local (normal) volatility

$$\sigma_F^2(F,t) = \mathbb{E}\left[\sigma_F^2(F,V) \mid F = K\right] = \frac{\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right]}{p(F,t)} \;.$$

Plugging the Kolmogorov equation (19) in equation (18) and integrating twice by part on the F variable, the Call price is finally obtained as an integral over time at strike K:

$$\mathrm{Call}(K,T) = e^{-rT}\left[(F_0 - K)^+ + \frac{1}{2}\int_0^T dt\,\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right]\right] \;. \tag{20}$$

Using expression (15) for the integrand, the integral over time can be computed:

$$\frac{1}{2}\int_0^T dt\,\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right] =$$

$$\frac{1}{\sqrt{2}}\,e^{-\widetilde{C}}\left[\sqrt{\frac{t}{\pi}}e^{-\frac{B}{t}} - \sqrt{B}\,\mathrm{erfc}\left(\sqrt{\frac{B}{t}}\right) - \frac{\widetilde{D}}{3}\left(\sqrt{\frac{t}{\pi}}(t-2B)e^{-\frac{B}{t}} + 2B^{3/2}\,\mathrm{erfc}\left(\sqrt{\frac{B}{t}}\right)\right)\right]$$

$$+\,o\left(t^{5/2}\,e^{-\frac{B}{t}}\right)$$

where erfc is the complementary error function, equal to the cumulative of the standard normal distribution up to $\sqrt{2}$ factors:

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{+\infty} dy\,e^{-y^2} = 2\mathcal{N}\left(-\sqrt{2}x\right) \;.$$

The asymptotic expansion of this function at $+\infty$

$$\mathrm{erfc}(x) = \frac{e^{-x^2}}{x\sqrt{\pi}}\left(1 - \frac{1}{2x^2} + \frac{3}{4x^4} + o\left(\frac{1}{x^4}\right)\right)$$

with $x = \sqrt{\frac{B}{t}}$ gives the asymptotic expansion of the time value:

$$\frac{1}{2}\int_0^T dt\,\mathbb{E}\left[\sigma_F^2(K,V)\delta(F-K)\right] = \frac{T^{3/2}}{2\sqrt{2\pi}}e^{-\frac{B}{T} - \widetilde{C} - \ln(B) - \widetilde{D}\,T - \frac{3}{2B}T} + o(T) \;. \tag{21}$$

3.4 Implied volatility

The final step consists in computing the same expansion for the Black–Scholes model, which is simpler as there is no stochastic volatility to be integrated. The metric is given by the inverse of the variance:

$$g_{FF} = \frac{1}{\sigma^2 F^2} \;.$$

The Christoffel symbol is therefore

$$\Gamma_{FF}^F = -\frac{1}{F} \;.$$

10

The $\mathbb{R}$-connection components are computed using (4) and (5):

$$A_F = \frac{1}{2F}$$

$$Q = \frac{\sigma^2}{8} .$$

The geodesic distance is

$$d(F_0, K) = \left| \int_{F_0}^{K} \frac{\mathrm{d}F}{\sigma F} \right| = \frac{1}{\sigma} \left| \ln \frac{K}{F_0} \right| .$$

The Van Vleck–Morette determinant is simply

$$\Delta(F_0, K) = 1 .$$

The parallel transport is

$$\mathcal{P}(F_0, K) = e^{-\int_{F_0}^{K} \mathrm{d}F \, A_F} = \sqrt{\frac{F}{K}}$$

Putting all these elements together, the heat kernel expansion of $p(K, t)$ is according to (9)

$$p(K, t) = \frac{1}{\sigma K \sqrt{2\pi t}} \sqrt{\frac{F}{K}} e^{-\frac{\ln^2 \frac{K}{F}}{2\sigma^2 t}} \left(1 - \frac{\sigma^2}{8} t + o(t) \right)$$

Multiplying by the local variance, we get

$$\mathbb{E}\left[\sigma_F^2(K) \delta(F - K) \right] = \sigma^2 K^2 p(K, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{B_{\mathrm{BS}}}{t} - \widetilde{C}_{\mathrm{BS}} - \widetilde{D}_{\mathrm{BS}} t + o(t)} \tag{22}$$

with

$$B_{\mathrm{BS}} = \frac{1}{2\sigma^2} \ln^2 \frac{K}{F_0}$$

$$\widetilde{C}_{\mathrm{BS}} = -\ln(\sigma) - \frac{1}{2} \ln(KF_0)$$

$$\widetilde{D}_{\mathrm{BS}} = \frac{\sigma^2}{8} .$$

(In fact formula (22) is exact: there is no $o(t)$ correction and it can be integrated exactly to get the Black–Scholes formula.)

Writing equation (21) for both the stochastic volatility model and the Black-Scholes model, the implied volatility is such that both quantities are equal:

$$\frac{B}{T} + \widetilde{C} + \ln(B) + \widetilde{D}T + \frac{3}{2B}T = \frac{B_{\mathrm{BS}}}{T} + \widetilde{C}_{\mathrm{BS}} + \ln(B_{\mathrm{BS}}) + \widetilde{D}_{\mathrm{BS}}T + \frac{3}{2B_{\mathrm{BS}}}T + o(T) . \tag{23}$$

Expanding the implied volatility σ as a Taylor expansion

$$\sigma(K, T) = \sigma_0(K) + \sigma_1(K)T + \sigma_2(K)T^2 + o(T^2)$$

and plugging this into equation (23) on the Black–Scholes side, we get

$$\frac{1}{2\sigma_0^2 T} \ln^2 \left(\frac{K}{F_0} \right) \left(1 - 2\frac{\sigma_1}{\sigma_0} T - 2\frac{\sigma_2}{\sigma_0} T^2 + 3 \left(\frac{\sigma_1}{\sigma_0} \right)^2 T^2 \right) - \ln(\sigma_0) - \frac{\sigma_1}{\sigma_0} T - \frac{1}{2} \ln(KF_0)$$

$$+ \ln \left(\frac{1}{2\sigma_0^2} \ln^2 \frac{K}{F_0} \right) - 2\frac{\sigma_1}{\sigma_0} T + \frac{\sigma_0^2}{8} T + \frac{3\sigma_0^2}{\ln^2 \frac{K}{F_0}} T = \frac{B}{T} + \widetilde{C} + \ln(B) + \widetilde{D} T + \frac{3}{2B} T + o(T) .$$

Coefficients must be equal at each order in T, which gives our final expansion of the implied volatility.

Power -1 gives the order 0 implied volatility

$$\sigma_0 = \frac{\left|\ln\frac{K}{F_0}\right|}{\sqrt{2B}} = \frac{\left|\ln\frac{K}{F_0}\right|}{d(F_0, V_0; K, V_{\min})} \tag{24}$$

which was already obtained in [BBF02] and [HL05].

The first order correction is extracted from the constant term:

$$\frac{\sigma_1}{\sigma_0} = -\frac{\widetilde{C} + \ln\left(\sigma_0\sqrt{KF_0}\right)}{2B} \ . \tag{25}$$

Finally the $O(T)$ term gives the second order correction:

$$\frac{\sigma_2}{\sigma_0} = \frac{3}{2}\left(\frac{\sigma_1}{\sigma_0}\right)^2 - \frac{1}{2B}\left(\widetilde{D} + 3\frac{\sigma_1}{\sigma_0} - \frac{\sigma_0^2}{8}\right) \ . \tag{26}$$

This gives our final result as the implied volatility expansion

$$\sigma = \sigma_0\left(1 + \frac{\sigma_1}{\sigma_0}T + \frac{\sigma_2}{\sigma_0}T^2 + o(T^2)\right) \ . \tag{27}$$

We stress that this result is exact in strike: for a given strike, we have computed exactly the three first coefficients of the Taylor expansion. Moreover, contrary to other expansions, the order 1 expansion is extracted from the order 0 expansion of the probability. This technique allows us to extract a second order term for the implied volatility from the order 1 term in the probability expansion. This method can be used to compute the Taylor expansion of implied volatility up to any order, although the computation becomes more complicated and involves integrals of increasing dimension: the a_k coefficient of the heat kernel expansion involves $k + 1$ integrals.

3.5 At the money

The computation we have performed makes the implicit hypothesis that we are not exactly at the money: $K \neq F_0$. Otherwise, the dominant term in the exponential would vanish and we could not use the asymptotic expansion of the erfc function at infinity. Precisely at the money, we should use instead a Taylor expansion in 0. As the implied volatility surface is smooth, we just take the limit of formulas (24), (25) and (26) at $K \to F_0$. If we perform instead the Taylor expansion of the erfc function at 0, we find only the two first orders

$$\sigma_0(F_0) = \frac{e^{-\widetilde{C}(F_0)}}{F_0}$$

$$\frac{\sigma_1}{\sigma_0}(F_0) = \frac{1}{3}\left(\frac{\sigma_0^2}{8} - \widetilde{D}(F_0)\right) \ .$$

Careful Taylor expansions of all quantities at the money can be used to check that this is indeed the limit of equations (24) and (25). Moreover, it can be seen that these limit are conditions for formulas (25) and (26) to be convergent, as B goes to 0 at the money (at order 2 in the geodesic distance, which means that the numerators must in fact vanish at order 2).

3.6 CEV volatility

Instead of Black volatility, the asymptotic expansion can be computed for other local volatility models. Without stochastic volatility, the SABR model reduces to the CEV model. The local volatility part of the model is thus taken into account exactly without introducing approximation besides the stochastic corrections. In view of our application to the SABR model, we will compute here a CEV implied volatility. There are closed formulas for this model, involving Bessel functions. This implied volatility can therefore be used in the CEV pricing formula in order to get the price of the option.

For a CEV model with parameter β_0 and volatility factor σ, such that

$$\mathrm{d}F = \sigma F^{\beta_0} \mathrm{d}W \ ,$$

the function B, $\widetilde{C}$ and $\widetilde{D}$ are

$$B_0 = \frac{1}{2\sigma^2} \ln^2(q_0)$$

$$\widetilde{C}_0 = -\ln(\sigma) - \frac{1}{2}\beta_0 \ln(KF_0)$$

$$\widetilde{D}_0 = \frac{\beta_0(2-\beta_0)\sigma^2}{8K^{1-\beta_0}F_0^{1-\beta_0}}$$

with

$$q_0 = \begin{cases} \dfrac{K^{1-\beta_0} - F_0^{1-\beta_0}}{1 - \beta_0} & \beta_0 < 1 \\[2ex] \ln\left(\dfrac{K}{F_0}\right) & \beta_0 = 1 \ . \end{cases}$$

Formulas (24), (25) and (26) are modified as follows.

$$\sigma_0 = \frac{|q_0|}{\sqrt{2B}} = \frac{|q_0|}{d(F_0, V_0; K, V_{\min})} \tag{28}$$

$$\frac{\sigma_1}{\sigma_0} = -\frac{\widetilde{C} + \ln(\sigma_0) + \frac{1}{2}\beta_0 \ln(KF_0)}{2B} \ . \tag{29}$$

$$\frac{\sigma_2}{\sigma_0} = \frac{3}{2}\left(\frac{\sigma_1}{\sigma_0}\right)^2 - \frac{1}{2B}\left(\widetilde{D} + 3\frac{\sigma_1}{\sigma_0} - \frac{\beta_0(2-\beta_0)\sigma_0^2}{8K^{1-\beta_0}F_0^{1-\beta_0}}\right) \ . \tag{30}$$

This gives the CEV implied volatility expansion

$$\sigma = \sigma_0 \left(1 + \frac{\sigma_1}{\sigma_0}T + \frac{\sigma_2}{\sigma_0}T^2 + o(T^2)\right) \ . \tag{31}$$

The Black implied volatility formulas correspond to the special case $\beta_0 = 1$. The Bachelier (*i.e.* normal) implied volatility would correspond to $\beta_0 = 0$.

3.7 Generalization

This technique can be generalized easily to other parameterizations of the options prices. Consider a model with local volatility or stochastic volatility, for which there are closed form formulas for European option prices. It can be used as a proxy in the following way.

- Denoting by z_i the parameters of the model, compute $B_*(z_i)$, $\widetilde{C}_*(z_i)$ and $\widetilde{D}_*(z_i)$, the quantities B, $\widetilde{C}$ and $\widetilde{D}$ of the asymptotic expansion (21) for this model at a given strike.

- Find parameters $z_i^{(0)}$ such that $B_*(z_i^{(0)}) = B$ (there can be several solutions).

- Choose a one-dimensional subset of the parameters $z_i = z_i(\lambda)$ which allows a wide range of option prices at the given strike and such that $z_i(0) = z_i^{(0)}$.

- Compute derivatives of $B_*(z_i)$, $\widetilde{C}_*(z_i)$ and $\widetilde{D}_*(z_i)$ with respect to λ at $\lambda = 0$. We use the notation $B_* = B_*(z_i(0))$, $B_*' = \partial_\lambda B_*(z_i(\lambda)) \mid_{\lambda=0} \dots$

- Write a Taylor expansion $\lambda(T) = \lambda_1 T + \lambda_2 T^2 + o(T^2)$ and write the equality of the asymptotic expansion (21) for the model and the proxy model:

$$
\frac{B}{T} + \widetilde{C} + \ln(B) + \widetilde{D}T + \frac{3}{2B}T = \frac{B_* + \lambda_1 T B_*' + \lambda_2 T^2 B_*' + \frac{1}{2}\lambda_1^2 T^2 B_*''}{T}
$$
$$
+ \widetilde{C}_* + \lambda_1 T \widetilde{C}_*' + \ln(B_* + \lambda_1 T B_*') + \widetilde{D}_* + \frac{3}{2B_*}T + o(T) \ .
$$

- This gives the Taylor expansion of λ:

$$
\lambda_1 = \frac{\widetilde{C} - \widetilde{C}_* + \ln(B) - \ln(B_*)}{B_*'}
$$
$$
\lambda_2 = \frac{\widetilde{D} - \widetilde{D}_* - \lambda_1 \frac{B_*'}{B_*} - \lambda_1 \widetilde{C}_* ' - \frac{1}{2}\lambda_1^2 B_*''}{B_*'} \ .
$$

- Plug parameters $z_i(\lambda_1 T + \lambda_2 T^2)$ into the closed form option price of the proxy model to get an approximate price of the option in the real model.

The closer the models are, the better the approximation is. It is clear that if the proxy model is the real model itself, there are no corrections at all. This procedure consists in approximating only the differences between models at a given strike and not the option price itself. In the basic case of section 3.4 where the proxy model is the Black-Scholes model, the approximation leverages on the fact that the volatility surface is more regular than the option price.

4   SABR Model

4.1   Model

The SABR Model [HKLW02] is a stochastic volatility model where the volatility is a local volatility function multiplied by a lognormal stochastic volatility:

$$
\begin{aligned}
\mathrm{d}F &= VC(F)\mathrm{d}W_1 \\
\mathrm{d}V &= \nu V \mathrm{d}W_2
\end{aligned}
\tag{32}
$$

with $\langle \mathrm{d}W_1 \mathrm{d}W_2 \rangle = \rho \mathrm{d}t$. The initial value for V is the parameter[4] α:

$$
\alpha = V(0) \ .
$$

---

[4]We use the standard notation of α for the initial value of the volatility variable in the SABR model instead of V_0 as in the previous section.

$C(F)$ is a local volatility function, which is generally

$$C(F) = F^\beta$$

β is a number between 0 and 1 which controls the local skew. 0 corresponds to a normal process and 1 to a lognormal process. The implied volatility at time 0 and at the money is the local volatility αF_0^β.

For $\beta \le \frac{1}{2}$, the origin $F = 0$ can be reached with finite probability in finite time. If F corresponds to a positive variable, a boundary condition must be imposed. The asymptotic expansion does not distinguish between different boundary conditions, as the computation is local around the geodesic path. It is no longer valid when the probability of reflection or absorbtion at the origin modifies the probability distribution at the strike considered in a significant way.

4.2   Order 0: metric

In order to compute the order 0 implied volatility, the only geometric object involved is the metric. According to the dictionary of section 2.2, its inverse is the covariance matrix

$$g^{ij} = \begin{pmatrix} V^2 C(F)^2 & \rho \nu V^2 C(F) \\ \rho \nu V^2 C(F) & \nu^2 V^2 \end{pmatrix}.$$

This matrix is first simplified by changing the variable F to

$$q = \int_{F_0}^{F} \frac{\mathrm{d}F}{C(F)} \tag{33}$$

which for $C(F) = F^\beta$ reads for $\beta \ne 1$

$$q = \frac{F^{1-\beta} - F_0^{1-\beta}}{1 - \beta}$$

and for $\beta = 1$

$$q = \ln\left(\frac{F}{F_0}\right).$$

In addition, we rescale the time such that ν disappears of the equations while keeping the same solution of the equations (the variances which are the physical quantities are not changed):

$$\begin{aligned} t &\longrightarrow \nu^2 t \\ \alpha &\longrightarrow \frac{\alpha}{\nu} \\ \nu &\longrightarrow 1 \end{aligned}$$

At the end of the computation, the inverse transformation must be applied to the implied volatility:

$$\sigma \nu \longleftarrow \sigma$$

The matrix in the set of variables (q, V) after this rescaling is

$$g^{ij} = V^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

This is diagonalized by going from variables (q, V) to (x, y) with

$$
\begin{aligned}
x &= \frac{q - \rho V}{\sqrt{1 - \rho^2}} \\
y &= V .
\end{aligned}
$$

The covariance matrix becomes

$$
g^{ij} = y^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .
$$

and its inverse is the metric

$$
g_{ij} = \frac{1}{y^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
$$

which corresponds to the infinitesimal distance

$$
ds^2 = \frac{dx^2 + dy^2}{y^2} .
$$

This geometry corresponds to the hyperbolic plane, in the Poincaré half-plane representation ($y > 0$) [HLW05, HL05]. Geodesics are vertical lines and semi-circles orthogonal to the $y = 0$ axis. The geodesic distance between two point (x_1, y_1) and (x_2, y_2) can be computed:

$$
d(x_1, y1; x_2, y2) = \cosh^{-1}\left(1 + \frac{(x_2 - x_1)^2 + (y_2 - y_1)^2}{2 y_1 y_2} \right) .
$$

In the (q, V) variables, going from $q = 0, V = \alpha$ to q, V the geodesic distance is

$$
d(0, \alpha; q, V) = \cosh^{-1}\left(1 + \frac{q^2 + (V - \alpha)^2 - 2\rho q(V - \alpha)}{2(1 - \rho^2)\alpha V} \right) .
$$

For a given strike, *i.e.* a given q, it is minimized by the volatility

$$
V_{\min} = \sqrt{\alpha^2 + 2\rho \alpha q + q^2}
$$

and the minimal distance is

$$
d(0, \alpha; q) = \cosh^{-1}\left(\frac{V_{\min} - \rho q - \rho^2 \alpha}{(1 - \rho^2)\alpha} \right) = \left| \ln\left(\frac{V_{\min} + \rho \alpha + q}{(1 + \rho)\alpha} \right) \right| .
$$

Equation (24) gives the order 0 implied volatility

$$
\sigma_0 = \frac{\ln\left(\dfrac{K}{F_0} \right)}{\ln\left(\dfrac{V_{\min} + \rho \alpha + q}{(1 + \rho)\alpha} \right)} .
$$

(We have dropped the absolute values as the numerator and the denominator have the same sign.)

Plugging the expression for $V_{\min}$ and going back to the original time, with ν factors, the order 0 implied volatility for the SABR model is

$$
\sigma_0 = \frac{\nu \, \ln\left(\dfrac{K}{F_0} \right)}{\ln\left(\dfrac{\sqrt{\alpha^2 + 2\rho \alpha \nu q + \nu^2 q^2} + \rho \alpha + q\nu}{(1 + \rho)\alpha} \right)} \tag{34}
$$

with

$$q = \begin{cases} \dfrac{K^{1-\beta} - F_0^{1-\beta}}{1-\beta} & \beta < 1 \\[2ex] \ln\left(\dfrac{K}{F_0}\right) & \beta = 1 \ . \end{cases}$$

At the money, the limit of this expression is simply

$$\sigma_0(F_0) = \alpha F_0^{\beta-1}$$

which is the local volatility.

4.3 Order 1: connection

To compute the order 1 correction, we need the scalar factor in the time value expansion, given by $\widetilde{C}$ in equation (16), with C given in equation (13).

For the hyperbolic plane, the Van Vleck–Morette determinant can be computed as a function of the geodesic distance:

$$\Delta = \frac{d}{\sinh(d)} \ .$$

We need also

$$\begin{aligned} \sigma_F(K, V) &= VC(K) = VK^\beta \\ g(K, V) &= \frac{1}{V^4 C(K)^2 (1-\rho^2)} = \frac{1}{V^4 K^{2\beta}(1-\rho^2)} \end{aligned}$$

Using that $d' = 0$ for the minimum distance at $V = V_{\min}$, we compute

$$B'' = dd'' = \frac{d}{\alpha V_{\min}(1-\rho^2)\sinh(d)}$$

Terms simplify against each other to give at $V = V_{\min}$

$$\widetilde{C} = -\ln\left(\sqrt{\alpha V_{\min}}K^\beta\right) + \mathcal{A} \tag{35}$$

$\mathcal{A}$ is the integral of the connection 1-form on the geodesic. According to formula (4), the connection is given by

$$A = \frac{1}{2(1-\rho^2)}\, d\ln(C(F)) - \frac{\rho C'(F)}{2(1-\rho^2)}\, dV \ . \tag{36}$$

In fact, A can be rewritten in (x, y) variables as

$$A = \frac{C'(F)}{2\sqrt{1-\rho^2}}\, dx \ .$$

It must be integrated from

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} \frac{-\rho\alpha}{\sqrt{1-\rho^2}} \\ \alpha \end{pmatrix}$$

to

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} \frac{q-\rho V}{\sqrt{1-\rho^2}} \\ V \end{pmatrix} \ .$$

If $\beta = 1$, the 1-form

$$A = \frac{1}{2\sqrt{1 - \rho^2}} \, dx$$

is exact and can be integrated directly:

$$\mathcal{A} = \frac{q - \rho V + \rho \alpha}{2(1 - \rho^2)} = \frac{1}{2} \ln\left(\frac{K}{F_0}\right) + \frac{\rho}{2(1 - \rho^2)} \left(\rho \ln\left(\frac{K}{F_0}\right) - V + \alpha\right) \, .$$

We consider now the general case where $\beta < 1$. If $x_1 = x_2$, the geodesic is vertical line. As A is along dx, its integral is zero:

$$\mathcal{A} = 0$$

In other cases, the first part of equation (36) is an exact form and can be integrated directly:

$$\int_{(F_0, \alpha)}^{(K, V)} \frac{1}{2(1 - \rho^2)} \, d\ln(C(F)) = \frac{1}{2(1 - \rho^2)} \ln\left(\frac{C(K)}{C(F)}\right) = \frac{\beta}{2(1 - \rho^2)} \ln\left(\frac{K}{F}\right) \, . \tag{37}$$

The second part must be integrated on the geodesic path. The geodesic is a semi-circle with origin $(X, 0)$, radius R and going through (x_1, x_2) and (y_1, y_2). The origin is therefore

$$X = \frac{x_2^2 - x_1^2 + y_2^2 - y_1^2}{2(x_2 - x_1)} \tag{38}$$

and the radius

$$R = \sqrt{y_1^2 + (x_1 - X)^2} \, . \tag{39}$$

We parameterize the geodesic by $t = \tan(\theta/2)$ where θ is the angle on the circle:

$$x = X + R \frac{1 - t^2}{1 + t^2}$$
$$y = R \frac{2t}{1 + t^2} \, .$$

In this parametrization, the geodesic distance is given by

$$ds = d\ln(t)$$

and we can compute

$$-\int_{\mathcal{C}} \frac{\rho \beta F^{\beta-1}}{2(1 - \rho^2)} dV = -\frac{\rho^2 \beta}{2(1 - \rho^2)} \ln\left(\frac{K}{F}\right) - \frac{\rho \beta}{(1 - \beta)\sqrt{1 - \rho^2}} [G(t_2) - G(t_1)] \tag{40}$$

with

$$G(t) = \tan^{-1}(t)$$

$$+ \begin{cases} -\dfrac{a + bX}{\sqrt{(a + bX)^2 - (1 - \beta)^2 R^2}} \tan^{-1}\left(\dfrac{cR + t(a + b(X - R))}{\sqrt{(a + bX)^2 - (1 - \beta)^2 R^2}}\right) & (a + bX)^2 > (1 - \beta)^2 R^2 \\ \dfrac{a + bX}{cR + t(a + b(X - R))} & (a + bX)^2 = (1 - \beta)^2 R^2 \\ \dfrac{a + bX}{\sqrt{(1 - \beta)^2 R^2 - (a + bX)^2}} \tanh^{-1}\left(\dfrac{cR + t(a + b(X - R))}{\sqrt{(1 - \beta)^2 R^2 - (a + bX)^2}}\right) & (a + bX)^2 < (1 - \beta)^2 R^2 \end{cases}$$

$$\tag{41}$$

$$a = F_0^{1-\beta}$$
$$b = (1-\beta)\sqrt{1-\rho^2}$$
$$c = (1-\beta)\rho$$

and

$$t_i = \sqrt{\frac{R - x_i + X}{R + x_i - X}} \ . \tag{42}$$

Summing equations (37) and (40), the integral of the connection is finally

$$\mathcal{A} = \frac{\beta}{2} \ln\left(\frac{K}{F}\right) - \frac{\rho\beta}{(1-\beta)\sqrt{1-\rho^2}} \left[G(t_2) - G(t_1)\right] \ . \tag{43}$$

Replacing $\mathcal{A}$ in equation (35) we get

$$\widetilde{C} = -\frac{1}{2}\ln\left(\alpha F_0^\beta V_{\min} K^\beta\right) + \begin{cases} -\dfrac{\rho\beta}{(1-\beta)\sqrt{1-\rho^2}} \left[G(t_2) - G(t_1)\right] & \beta < 1 \\[2ex] \dfrac{\rho}{2(1-\rho^2)} \left(\rho\ln\left(\dfrac{K}{F_0}\right) - V_{\min} + \alpha\right) & \beta = 1 \end{cases} \tag{44}$$

Restoring the factor ν, the order 1 correction is given by equation (25):

$$\frac{\sigma_1}{\sigma_0} = -\nu^2 \frac{\widetilde{C} + \ln\left(\dfrac{\sigma_0}{\nu}\sqrt{KF_0}\right)}{\ln^2\left(\dfrac{\sqrt{\alpha^2 + 2\rho\alpha\nu q + \nu^2 q^2} + \rho\alpha + q\nu}{(1+\rho)\alpha}\right)} \tag{45}$$

σ_0 is given by equation (34), $\widetilde{C}$ by equation (44) (with α divided by ν, also inside $V_{\min}$), where X, R, t_1 and t_2 are given in equations (38), (39) and (42) from

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} \dfrac{-\rho\alpha}{\nu\sqrt{1-\rho^2}} \\ \dfrac{\alpha}{\nu} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} \dfrac{\nu q - \rho\sqrt{\alpha^2 + 2\rho\alpha\nu q + \nu^2 q^2}}{\nu\sqrt{1-\rho^2}} \\ \dfrac{\sqrt{\alpha^2 + 2\rho\alpha\nu q + \nu^2 q^2}}{\nu} \end{pmatrix}$$

and $G(t)$ is defined in formula (41).

Using this expression, the first order implied volatility is

$$\sigma = \sigma_0 \left(1 + \frac{\sigma_1}{\sigma_0} t + o(t)\right)$$

which is valid for all positive strikes. Exactly at the money, the formula we give must be replaced by its limit, which can be computed by a Taylor expansion or numerically. At the money and only at the money it appears to be equal to the original HKLW formula. This is not surprising as their expansion is in fact an expansion in both maturity and moneyness.

$$\frac{\sigma_1}{\sigma_0}(F_0) = \frac{1}{24}\alpha^2(1-\beta)^2 F_0^{2(\beta-1)} + \frac{1}{4}\rho\alpha\nu\beta F_0^{\beta-1} + \frac{1}{12}\nu^2 - \frac{1}{8}\rho^2\nu^2 \ . \tag{46}$$

4.4 Order 2

To compute the second order correction to implied volatility, we need to compute $\widetilde{D}$ as defined in equation (17), with $D = -a_1$ defined in equation (10).

We have to compute a_1 as defined in equation (10). Most of the integration can be done analytically. We have first the integral of Q along the geodesic:

$$a_1^{(Q)} = -\frac{1}{d}\int_{\mathcal{C}} Q \mathrm{d}s \ .$$

According to equation (5), Q is

$$Q = \frac{\beta}{4}\left(1 - \beta + \frac{\beta}{2(1-\rho^2)}\right)\frac{V^2}{F^{2(1-\beta)}} \ .$$

Using the values defined in the previous section for X, R, t_1, t_2, a, b and c, its integral along the geodesic is

$$a_1^{(Q)} = \frac{\beta}{2}\left(1 - \beta + \frac{\beta}{2(1-\rho^2)}\right)\frac{R^2}{(1-\beta)^2 R^2 - (a+bX)^2}\frac{H(t_2)-H(t_1)}{\ln(t_2)-\ln(t_1)} \tag{47}$$

with

$$H(t) = \frac{a + b(R+X) + cRt}{(a+bX)(1+t^2) + bR(1-t^2) + 2cRt}$$

$$+ \begin{cases} \dfrac{cR}{\sqrt{(a+bX)^2 - (1-\beta)^2 R^2}}\tan^{-1}\left(\dfrac{cR + t(a+b(X-R))}{\sqrt{(a+bX)^2 - (1-\beta)^2 R^2}}\right) & (a+bX)^2 > (1-\beta)^2 R^2 \\[3ex] -\dfrac{cR}{\sqrt{(1-\beta)^2 R^2 - (a+bX)^2}}\tanh^{-1}\left(\dfrac{cR + t(a+b(X-R))}{\sqrt{(1-\beta)^2 R^2 - (a+bX)^2}}\right) & (a+bX)^2 < (1-\beta)^2 R^2 \ . \end{cases}$$

Note that in the denominator, the quantity $\ln(t_2) - \ln(t_1)$ is up to a sign the geodesic distance d. If $\beta = 1$, $a_1^{(Q)}$ reduces to

$$a_1^{(Q)} = -\frac{R\mid x_2 - x_1 \mid}{8(1-\rho^2)d} \ . \tag{48}$$

The Laplacian on the hyperbolic plane is in (x,y) coordinates

$$D^i D_i = y^2(\partial_x^2 + \partial_y^2) \ .$$

As the Van Vleck–Morette determinant $\Delta = \frac{s}{\sinh(s)}$ depends only on the geodesic distance s, its derivative on the orthogonal coordinate vanishes: $\partial_\perp\Delta = 0$. On the other hand, by definition the parallel transport on the geodesic curve has no covariant derivative along the curve: $\nabla_s\mathcal{P} = 0$. As a consequence, there is no crossed term and both terms decouple: we have

$$\mathcal{P}^{-1}\Delta^{-1/2}\nabla^i\nabla_i(\mathcal{P}\Delta^{1/2}) = \mathcal{P}^{-1}\nabla^i\nabla_i\mathcal{P} + \Delta^{-1/2}D^i D_i\Delta^{1/2}$$

(the $\mathbb{R}$ charge is carried only by $\mathcal{P}$).

The metric part can be integrated analytically [HLW05, Avr07]:

$$a_1^{(R)} = -\frac{1}{8}\left[1 + \frac{1}{d}\left(\frac{\cosh(d)}{\sinh(d)} - \frac{1}{d}\right)\right] \ . \tag{49}$$

The last part to integrate is the $\mathbb{R}$ connection term $\mathcal{P}^{-1}\nabla^i\nabla_i\mathcal{P}$. As the action of gauge transformations on the heat kernel expansion is fully carried by the parallel transport term $\mathcal{P}$, a_1 can only

depend on gauge-invariant quantities constructed from $F = dA$. We split therefore $A = A^{(0)} + A^{(1)}$ into a pure gauge part $A^{(0)}$ and $A^{(1)}$ such that $F = dA^{(1)}$:

$$A^{(0)} = \frac{1}{2} d \ln(C(F))$$

$$A^{(1)} = \frac{\rho^2}{2(1-\rho^2)} d \ln(C(F)) - \frac{\rho C'(F)}{2(1-\rho^2)} dV .$$

Forgetting $A^{(0)}$ which is pure gauge (it can be checked by hand that $A^{(0)}$ will not contribute), we denote by $\mathcal{P}^{(1)}$ the $A^{(1)}$ part of the parallel transport:

$$\mathcal{P}^{(1)} = e^{-\mathcal{A}^{(1)}} = e^{-\int_C A^{(1)}}$$

with

$$\mathcal{A}^{(1)} = \begin{cases} -\dfrac{\rho\beta}{(1-\beta)\sqrt{1-\rho^2}} [G(t_2) - G(t_1)] & \beta < 1 \\ \dfrac{\rho}{2(1-\rho^2)} \left(\rho \ln\left(\dfrac{K}{F_0}\right) - V + \alpha \right) & \beta = 1 . \end{cases} \tag{50}$$

Using $\mathcal{A}^{(1)}$ and $A^{(1)}$ which in (x, y) variables is

$$A_1 = \frac{\rho C'(F)}{2} \left(\frac{\rho}{\sqrt{1-\rho^2}} dx - dy \right) ,$$

we can rewrite

$$\frac{1}{2} \mathcal{P}^{-1} \nabla^i \nabla_i \mathcal{P} = \frac{1}{2} y^2 \left[-\left(\partial_x^2 + \partial_y^2\right) \mathcal{A}^{(1)} + \left(\partial_x \mathcal{A}^{(1)} - A_x^{(1)}\right)^2 + \left(\partial_y \mathcal{A}^{(1)} - A_y^{(1)}\right)^2 + \partial_x A_x^{(1)} + \partial_y A_y^{(1)} \right]$$

where the specific form of $A^{(1)}$ can be used to simplify terms

$$\partial_x A_x^{(1)} + \partial_y A_y^{(1)} = 0 .$$

Computing $\mathcal{A}^{(1)}$ and $A^{(1)}$ analytically, we compute numerically the x and y derivatives and integrate numerically along the geodesic curve to get

$$a_1^{(A)} = \frac{1}{2d} \int_C ds \, y^2 \left[-\left(\partial_x^2 + \partial_y^2\right) \mathcal{A}^{(1)} + \left(\partial_x \mathcal{A}^{(1)} - A_x^{(1)}\right)^2 + \left(\partial_y \mathcal{A}^{(1)} - A_y^{(1)}\right)^2 \right] . \tag{51}$$

This integral can be computed by a numerical quadrature with few points. For $\beta = 1$ the connection has in fact no curvature and therefore $a_1^{(A)} = 0$.

We compute also the following quantities (at $V = V_{\min}$):

$$d'' = \frac{1}{\alpha(1-\rho^2)V_{\min}\sinh(d)}$$

$$B'' = dd''$$

$$\frac{B^{(3)}}{B''} = -\frac{3}{V_{\min}} \tag{52}$$

$$\frac{B^{(4)}}{B''} = \frac{12}{V_{\min}^2} - 3d'' \left(\frac{\cosh(d)}{\sinh(d)} - \frac{1}{d} \right) .$$

We need finally $\widetilde{C}'$ and $\widetilde{C}''$. Using our decomposition $\mathcal{A} = \frac{\beta}{2} \ln \frac{K}{F} + \mathcal{A}^{(1)}$, we can write from formulas (13) and (16)

$$\widetilde{C} = -\frac{\beta}{2} \ln(KF) - \frac{1}{2} \ln\left(\frac{d}{\sinh(d)} \right) + \frac{1}{2} \ln(B'') + \ln(1-\rho^2) + \mathcal{A}_1 .$$

Its first and second derivatives in V at $V = V_{\min}$ are

$$\widetilde{C}' = \frac{1}{2}\frac{B^{(3)}}{B''} + \mathcal{A}^{(1)'}$$

$$\widetilde{C}'' = \frac{1}{2}d''\left(\frac{\cosh(d)}{\sinh(d)} - \frac{1}{d}\right) + \frac{1}{2}\frac{B^{(4)}}{B''} - \frac{1}{2}\left(\frac{B^{(3)}}{B''}\right)^2 + \mathcal{A}^{(1)''}.$$

We choose to differentiate $\mathcal{A}^{(1)}$ numerically, by finite difference, as the analytical expression is very long and hard to simplify.

Simplifying $\frac{\widetilde{C}''}{2B''}$ against $-a_1^{(R)}$ we get finally

$$\widetilde{D} = -a_1^{(Q)} - a_1^{(A)} + \frac{1}{8} + \frac{1}{2B''}\left[\mathcal{A}^{(1)''} - \mathcal{A}^{(1)'2} + \frac{3}{V_{\min}}\mathcal{A}^{(1)'} - \frac{3}{4V_{\min}^2}\right]$$

with $a_1^{(Q)}$ given in equation (47), $a_1^{(A)}$ in (51), B'' in (52) and $\mathcal{A}^{(1)}$ in (50) with $G(t)$ defined in equation (41). For $\beta = 1$, this expression can be simplified using equation (48) for $a_1^{(Q)}$ and

$$a_1^{(A)} = 0$$

$$\mathcal{A}^{(1)'} = -\frac{\rho}{2(1 - \rho^2)}$$

$$\mathcal{A}^{(1)''} = 0 .$$

We have finally obtained the second order corrective term

$$\frac{\sigma_2}{\sigma_0} = \frac{3}{2}\left(\frac{\sigma_1}{\sigma_0}\right)^2 - \frac{1}{2B}\left(\widetilde{D} + 3\frac{\sigma_1}{\sigma_0} - \frac{\sigma_0^2}{8}\right)$$

such that we get the quadratic approximation to implied volatility

$$\sigma = \sigma_0\left(1 + \frac{\sigma_1}{\sigma_0}T + \frac{\sigma_2}{\sigma_0}T^2 + o(T^2)\right) .$$

The computation has been done in redefined variables such that $\nu = 1$. To restore the ν factors, α must be replaced by $\frac{\alpha}{\nu}$, T by $\nu^2 T$ and the final implied volatility must be multiplied by ν.

At the money, the formula for $\frac{\sigma_2}{\sigma_0}$ looks divergent but its limit is well defined. We compute this limit numerically, although it could be done analytically.

4.5   CEV volatility

The results of section 3.6 can be used to invert he SABR volatility into a CEV fractional volatility. Using formulas of section 3.6 the implied CEV volatility is computed and used in the closed-form option prices of the CEV model.

This proves to be useful for low β at low strikes or with very small volatility of volatility: only the corrections to the CEV model which come from the stochastic volatility are approximated, not the local volatility part. For example, at the money the first order coefficient of the Black volatility which is given by equation 46 becomes for the CEV volatility

$$\frac{\sigma_1}{\sigma_0}(F_0) = \frac{1}{4}\rho\alpha\nu\beta F_0^{\beta-1} + \frac{1}{12}\nu^2 - \frac{1}{8}\rho^2\nu^2 .$$

The corrective term in $\alpha^2 T$ has disappeared.

4.6 Numerical results

We present in figures 1 and 2 the implied volatility given by our expansion and compare it to the implied volatility computed by a two-dimensional finite difference method scheme. We also show for comparison the implied volatility given by the original formula of [HKLW02]. In this example, parameters are $F_0 = 4$, $\alpha = 30\%$, $\beta = .7$, $\nu = 40\%$, $\rho = -.5$ The FDM scheme is a second order Yanenko scheme with 400 points in strike, 200 points in volatility and 30 time steps.

At very short maturities, all expansions are good. At first order our expansion is equal to the HKLW at the money but is more regular in strikes and is better in the wings as our computation does not involve any approximation in the moneyness. Our second order expansion is one order of magnitude more precise[5]. When maturity grows, first order expansions lose precision but the second order remain relatively good up to 10 years, where $\nu^2 T = 1.6$. At higher maturities, the second order expansion explodes quadratically and finally gives even negative volatilities at very long maturity and low strikes. At long maturities, a FDM or an other numerical method must be used, unless a valid long maturity expansion could be found.

5 Conclusion

We have presented a general method to compute a Taylor expansion in maturity of implied volatility of stochastic volatility models. We give exact formulas for the first and second order corrections. As an application, we have computed this expansion for the SABR model and compared it to the implied volatility given by a numerical scheme. It appears that it gives more precise results than the usual formula and extend the domain where a short maturity expansion can be used. Outside this range of validity, other methods must be used: numerical schemes or possibly other approximations.

If a closer model with closed formulas than Black-Scholes exists, we provide a method to use this model as a proxy to extend the domain of validity of the expansion. It would be interesting to see the results of this method for the SABR model with a stochastic volatility model as a proxy.
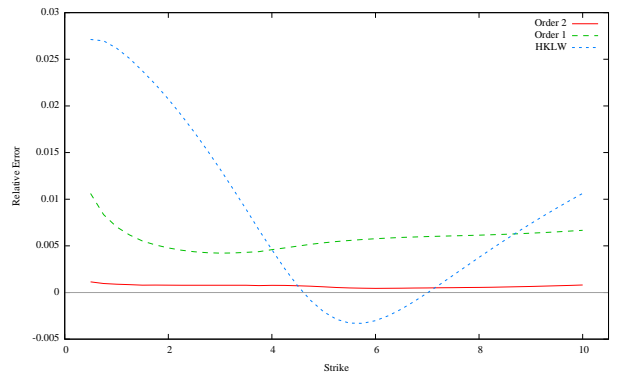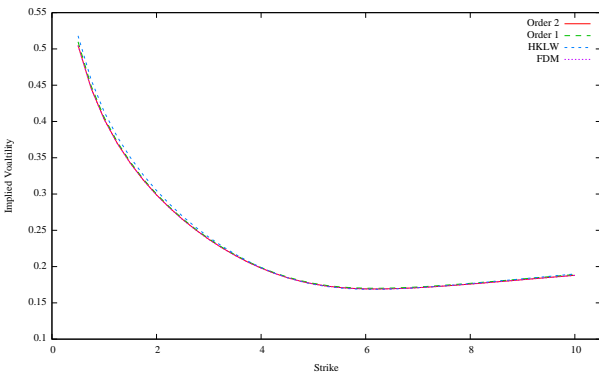
Obtaining exact option prices at all maturities would be a non-perturbative computation, which is a longstanding issue in theoretical physics.
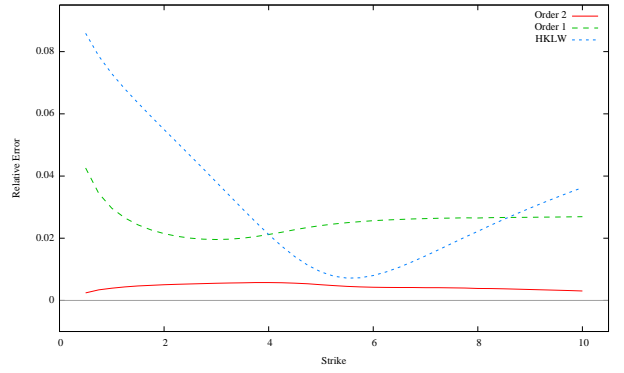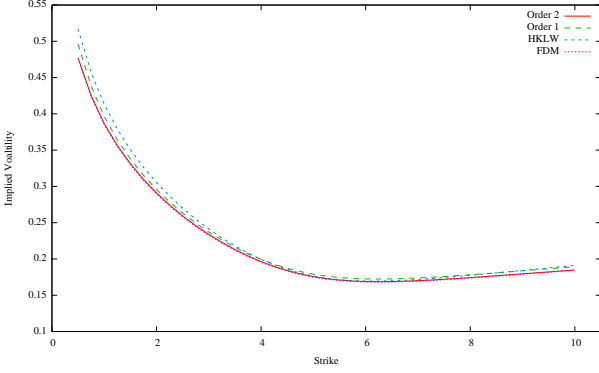
Acknowledgements

---

[5]In fact at even shorter maturities, the FDM scheme we use is less precise than this second order expansion, especially in the wings where the probability density is very small.
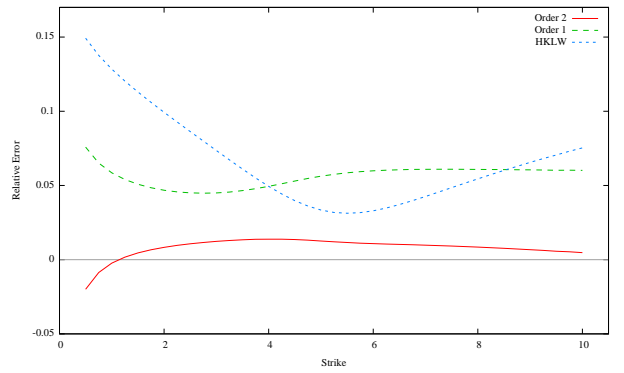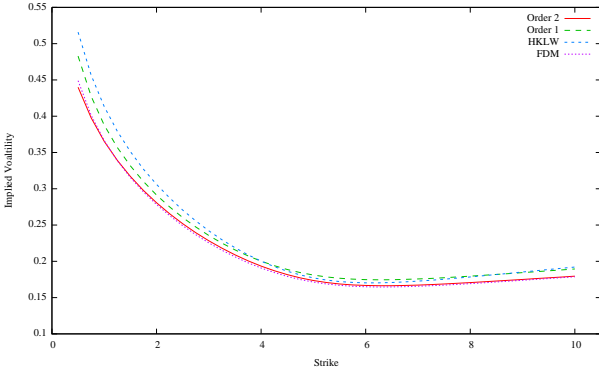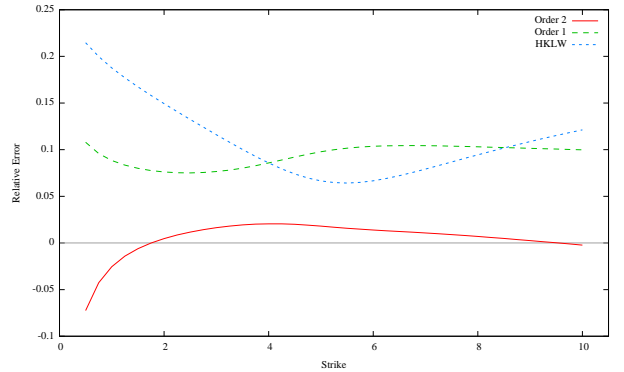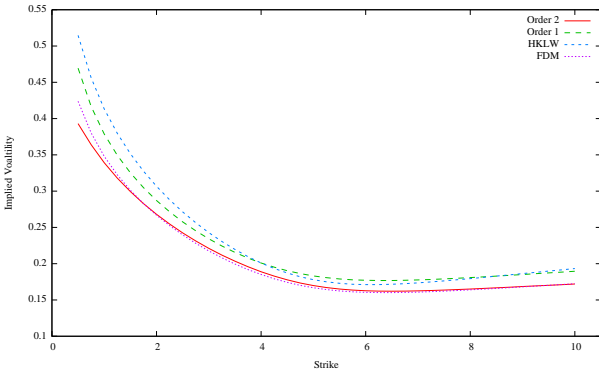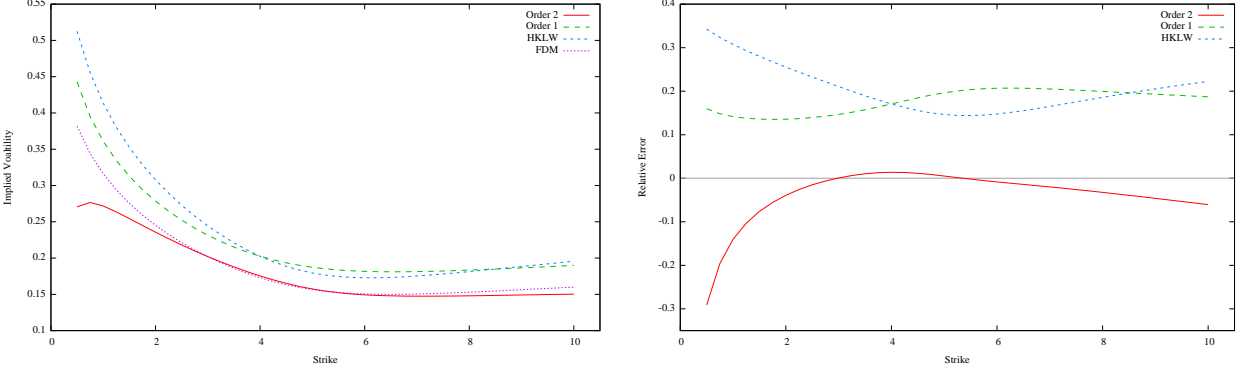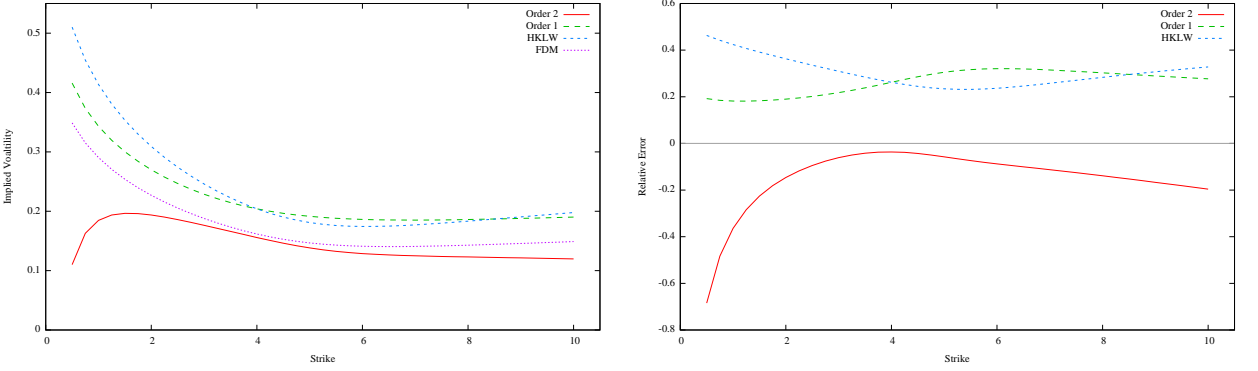
Figure 1: *Implied volatility and relative error for the SABR model with parameters $F_0 = 4$, $\alpha = 30\%$, $\beta = .7$, $\nu = 40\%$, $\rho = -.5$ and maturities 2.5 yr, 5 yr, 7.5 yr and 10 yr. On the left, implied volatility are plotted for our first order and second order expansions, the original formula of [HKLW02] and are compared to the result of a FDM solution. On the right are the relative errors with respect to this reference solution.*
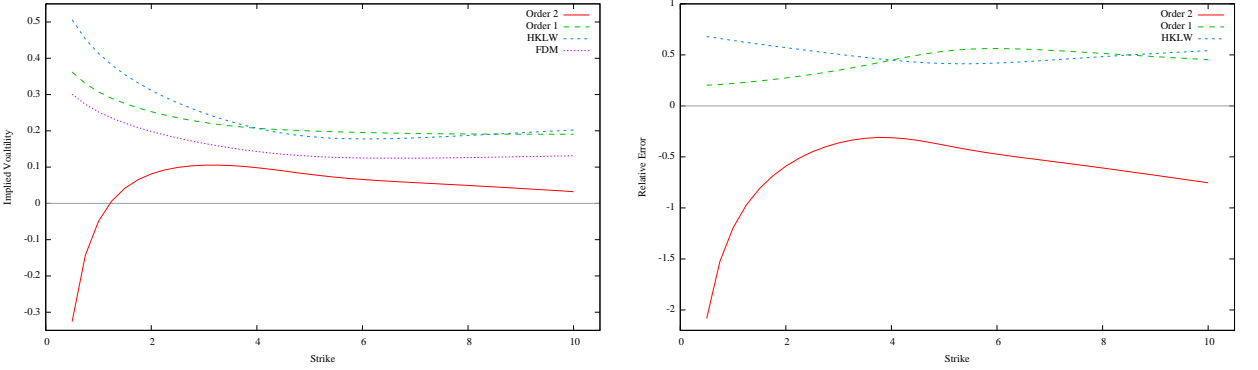
Figure 2: *Implied volatility and relative error for the SABR model with parameters $F_0 = 4$, $\alpha = 30\%$, $\beta = .7$, $\nu = 40\%$, $\rho = -.5$ and maturities 10 yr, 20 yr and 30 yr. On the left, implied volatility are plotted for our first order and second order expansions, the original formula of [HKLW02] and are compared to the result of a FDM solution. On the right are the relative errors with respect to this reference solution.*

Appendix A Mean reversion

At long maturity, the SABR model is not realistic as the volatility process is a geometric Brownian motion. In particular the variance of the volatility increases linearly in time. A direct extension would be to add mean reversion to the volatility process, either on the volatility or on the variance. The asymptotic expansion can still be computed at order 2. However its domain of validity is usually reduced: in addition to other conditions, the maturity must be small in front of the mean reversion characteristic time.

We impose mean reversion on the volatility process as

$$\mathrm{d}V = \nu V \mathrm{d}W_2 + \kappa(\overline{V} - V)\mathrm{d}t$$

The metric is not modified as it describes the diffusion part. Expressions for A and Q are modified as follows:

$$A = A_{\mathrm{SABR}} - \frac{\rho\kappa(V - \overline{V})}{\nu V^2(1 - \rho^2)F^\beta}\mathrm{d}F + \frac{\kappa(V - \overline{V})}{\nu^2 V^2(1 - \rho^2)}\mathrm{d}V$$

$$Q = Q_{\mathrm{SABR}} + \frac{1}{2}\frac{\kappa^2(V - \overline{V})^2}{\nu^2 V^2(1 - \rho^2)} + \frac{1}{2}\kappa - \kappa\frac{\overline{V}}{V} - \frac{1}{2}\frac{\rho\beta\kappa(V - \overline{V})}{\nu(1 - \rho^2)F^{1-\beta}} .$$

At first order, the integral of A along the geodesic is needed. Using the same notations as in section 4.3 where variables have been rescaled such that $\nu = 1$, it gets an additional term

$$\mathcal{A} = \mathcal{A}_{\mathrm{SABR}} + \frac{\rho\kappa}{\sqrt{1 - \rho^2}}\left[2\left(\tan^{-1}(t_2) - \tan^{-1}(t_1)\right) - \frac{\overline{V}}{R}\ln\left(\frac{t_2}{t_1}\right)\right] + \kappa\left[\ln\left(\frac{V}{\alpha}\right) + \frac{\overline{V}}{V} - \frac{\overline{V}}{\alpha}\right] .$$

The second order correction involves a one-dimensional integral which can be computed numerically by a quadrature with a few points, although it may be possible to get an analytical expression.

References

[Avr07]   Ivan G. Avramidi. Analytical and geometric methods for heat kernel applications in finance. `http://infohost.nmt.edu/~iavramid/notes/hkt/hktslides7b.pdf`, 2007.

[BBF02]   Henri Berestycki, Jérôme Busca, and Igor Florent. Asymptotics and calibration of local volatility models. *Quantitative finance*, 2(1):61–69, 2002.

[BBF04]   Henri Berestycki, Jérôme Busca, and Igor Florent. Computing the implied volatility in stochastic volatility models. *Communications on Pure and Applied Mathematics*, 57(10):1352–1373, 2004.

[Cox75]   John C. Cox. Notes on option pricing I: Constant elasticity of diffusions. Working Paper, 1975.

[DeW65]   Bryce S. DeWitt. *Dynamical Theory of Groups and Fields*. Gordon and Breach, 1965.

[Hes93]   Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[HKLW02] Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward. Managing smile risk. *Wilmott magazine*, 1(8):84–108, 2002. `http://ssrn.com/abstract=242367`.

[HL05]   Pierre Henry-Labordère. A general asymptotic implied volatility for stochastic volatility models. `http://arxiv.org/abs/cond-mat/0504317`, 2005.

[HL08]   Pierre Henry-Labordère. *Analysis, Geometry, and Modeling in Finance: Advanced Methods in Option Pricing*. Chapman & Hall/CRC, 2008.

[HLW05]   Patrick S. Hagan, Andrew S. Lesniewski, and Diana E. Woodward. Probability distribution in the SABR model of stochastic volatility. `http://www.lesniewski.us/papers/working/ProbDistrForSABR.pdf`, 2005.

[Var67a]   S.R.S. Varadhan. Diffusion processes in a small time interval. *Communications on Pure and Applied Mathematics*, 20(4):659–685, 1967.

[Var67b]   S.R.S. Varadhan. On the behavior of the fundamental solution of the heat equation with variable coefficients. *Communications on Pure and Applied Mathematics*, 20(2):431–455, 1967.

[Vas03]   Dmitri V. Vassilevich. Heat kernel expansion: User's manual. *Physics Report*, 388:279–360, 2003. `http://arxiv.org/abs/hep-th/0306138`.