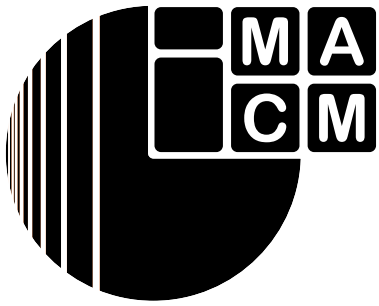


Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 20/01

Mike Felpel, Jörg Kienitz and Thomas A. McWalter

Effective Stochastic Volatility: Applications to ZABR-type Models

January 2020

http://www.math.uni-wuppertal.de

Effective Stochastic Volatility – Applications to ZABR-type Models –

Mike Felpel
Fachbereich Mathematik und Naturwissenschaften, Bergische Universtät Wuppertal
(mike.felpel@uni-wuppertal.de)

Jörg Kienitz
African Institute of Financial Markets and Risk Management, University of Cape Town
Fachbereich Mathematik und Naturwissenschaften, Bergische Universtät Wuppertal
(joerg.kienitz@math.uni-wuppertal.de)

Thomas A. McWalter
African Institute of Financial Markets and Risk Management, University of Cape Town
Department of Statistics, University of Johannesburg
(tom@analytical.co.za)

January 12, 2020

Abstract

There are numerous models for specifying the uncertainty of future instantaneous volatility or variance, including the Heston, SABR and ZABR models. Often it is observed that a specific stochastic volatility model is chosen not for particular dynamical features, relevant for exotic payoff structures, but instead for convenience and ease of implementation. The SABR model, with its semi-closed form approximate solution for the prices of vanilla options, is a well-known example. In this article, we consider a general approach that includes all practically relevant stochastic volatility models and introduces new variants of the ZABR model. In particular, we consider the mean-reverting ZABR and free ZABR models. We use the method of deriving an effective partial differential equation for the density. This approach leads to the known approximation formula for the SABR model, but also provides expressions for arbitrage-free models. Numerical experiments illustrate our approach.

Keywords: Stochastic volatility, SABR, ZABR, Free boundary ZABR, Mean-reverting ZABR, Effective PDE, Approximation formula

Contents

1 Introduction

This paper is concerned with modeling the implied volatility surface for an underlying asset. In our exposition, we consider the dynamics of the forward rate, which simplifies the SDEs considered by ensuring that the asset has no drift. The implied volatility surface is defined in relation to a reference valuation model with an analytic solution for vanilla European call and put options. For the underlying asset, F, strike value, K, and maturity time, T, the contract pays

$$\begin{aligned} V_T &= \max(F_T - K, 0) \text{ for a call,} \qquad \text{and} \\ V_T &= \max(K - F_T, 0) \text{ for a put.} \end{aligned}$$

Given the maturity, strike and current (time-t_0) price of the underlying, f, the reference formula used to compute the current price, V_{t_0}, must have only one additional degree of freedom, the *implied volatility*, which is specified in order to match the quoted market price, V_{Market}. The standard reference models used are the Black and Bachelier models. Since we shall only consider examples for interest rates, we use the Bachelier model, also known as the *Gauss* or *normal* model.

The Bachelier pricing formulas for calls and puts are

$$\begin{aligned} C_\phi(T, K, \sigma) &= (f - K)\,\Phi(d) + \sigma\sqrt{T}\phi(d) \qquad \text{and} \\ P_\phi(T, K, \sigma) &= (K - f)\,\Phi(-d) + \sigma\sqrt{T}\phi(d) \end{aligned}$$

with

$$d = \frac{f - K}{\sigma\sqrt{T - t_0}},$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the normal CDF and PDF respectively. Here, σ is called the *Bachelier*, *normal* or *bp* implied volatility. The choice of the reference model depends on the given market. For instance the Bachelier model can be applied to negative asset values (applicable to interest rates) and does not require the values to be bounded.

Often we are not only interested in a single option price, but require prices for a set of maturities and strikes. Consider a set $\mathcal{T} := \{T_1, T_2, \ldots, T_N\}$ of option maturities and let $\mathcal{K} := \{\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_N\}$, $\mathcal{K}_i := \{K_{i,1}, K_{i,2}, \ldots, K_{i,M_i}\}$ be sets of strike values indexed by the number of maturities considered. Usually $\mathcal{K}_i = \mathcal{K}_j$ for all $1 \le i, j \le N$. Now, consider the implied volatility for each quoted option with respect to $T_i \in \mathcal{T}$, $K_j \in \mathcal{K}$:

$$\begin{aligned} \Sigma_d : \mathcal{T} \times \mathcal{K} &\to \mathbb{R}^+ \\ (T, K) &\mapsto \sigma_d. \end{aligned}$$

The map Σ_d is called the discrete implied volatility surface. From this set, values for strikes $K \notin \mathcal{K}$ may be inferred by interpolation and extrapolation. The latter techniques must respect arbitrage relationships, requiring practitioners to use interpolation methods consistent with an arbitrage-free model. Furthermore, starting with the current levels of volatility as an input, it is relevant to consider the dynamics of the implied volatilities in order to manage exotic options. Such options may, for instance, include payoffs that strongly depend on the forward volatility. Examples include forward-starting options and cliquet options. Practitioners refer to the implied volatility surface as the skew or smile. These names originate from the shapes that the surfaces exhibit in typical market environments.

While there are many approaches for modeling the dynamics of instantaneous volatility, including Levy and jump-diffusion models, see, e.g., [23–25, 31] for a non-exhaustive list, practitioners most often use stochastic volatility models to manage this type of risk. Selecting a stochastic volatility model determines the continuous implied volatility surface. The process of matching to the observed discrete implied volatility surface is called calibration, and, once a model is calibrated, the continuous implied volatility surface may be used for interpolation and extrapolation. We consider the continuous implied volatility surface given by the map

$$\begin{aligned} \Sigma_{c,0} : [0,T] \times [K_l, K_u] &\to \mathbb{R}^+ \\ (T,K) &\mapsto \sigma_c \end{aligned}$$

and its dynamics $\Sigma_{c,t}(T,K)$, $t \in \mathbb{R}^+$. By choosing a stochastic volatility model these dynamics are implicitly determined by the model. In particular, we consider the coupled SDEs given by

$$\begin{cases} d\tilde{F}_t = C(\tilde{F}_t) v_t \, dW_t^{(1)}, & \tilde{F}_{t_0} = f, \\ dv_t = \mu(v_t)\, dt + \nu(v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases} \tag{1.1}$$

The choice of model and parameters should ensure the best fit to the current (discrete) market implied volatility surface and that the dynamics are suitable for risk management and trading of exotic contracts. Often ease of implementation determines the choice of the model, rather than model suitability. We provide a general modeling approach with a tractable computational framework that does not require this compromise.

To achieve numerical tractability, we use singular perturbation methods to derive an approximate PDE, called the *effective PDE*, for the marginal probability density

of the asset. Here, this probability density should be understood as

$$\mathbb{P}\left[F < \tilde{F}_t < F + dF \,\middle|\, \tilde{F}_{t_0} = f, v_{t_0} = \alpha\right].$$

This technique was originally introduced by Hagan *et al.* [9, 11, 14, 15] for SABR models. Our general framework provides an extension to this approach and covers most well known stochastic volatility models, including the SABR models mentioned above, but also the Stein-Stein [33] and Schoebel-Zhu [30] models.

Having derived the general effective equation, we are in a position to consider many stochastic volatility model variants. Naturally, our approach includes the standard SABR model as a special case, but it also includes the displaced SABR, mean-reverting SABR (mrSABR) and free boundary SABR (fSABR) models. Free boundary models have a volatility function given by $|F_t|^\beta$, and were introduced by Antonov *et al.* [4]. We also focus on ZABR type models, originally introduced by Andreasen and Huge [1]. We take this as a starting point and consider the displaced ZABR, mean-reverting ZABR (mrZABR) and free boundary ZABR (fZABR) models.

Our methodology relies on efficiently solving the one-dimensional effective PDE. The alternative would be to solve the two-dimensional problem. From a numerical perspective, a non-zero correlation between the asset and the volatility driver makes it impossible to apply alternating direction implicit (ADI) methods directly. For an approach that uses transforms and an application to the Heston model, see [17]. Transformations may be problematic when it comes to boundary conditions. The boundaries of the transformed dynamics may be more complicated than the original ones. Another approach is to handle the terms involving both quantities F_T and v_T by an explicit step, but this requires small time steps in order to avoid numerical instability. However, for the two-dimensional setting, the Peaceman-Rachford-Douglas method [6, 27] may be used; it may even be combined with an appropriate solver so that an explicit step is not necessary. Iterative solver algorithms, such as BiCGStab [7, 29], do not even require an explicit matrix representation. When considering examples to illustrate the new models, we use the numerical methods described in [13, 20, 21].

The rest of the paper proceeds as follows: In Section 2 we introduce the methodology and derive the effective PDE for the general framework. The whole derivation—including all details—can be found in Appendix A. In Subsection 2.2 we consider the approximating formula and then provide more detail on concrete models in Subsection 2.3. In Subsection 2.3.1 we use the method to analyze the smile dynamics of the ZABR model, while in Subsection 2.3.2 we investigate the extended ZABR models.

The approach taken to demonstrate numerical examples in Section 3 is based on one dimensional PDE solvers. We use numerical methods to compare the resulting stochastic volatility modeling approaches in Subsection 3.1 and the calibration behavior in Subsection 3.2. Finally, Subsection 3.3 compares our approach to a classical Monte Carlo approach [22]. Section 4 concludes with a summary, draws conclusions and outlines directions for future research.

2 Main Results

To provide a tractable way to compute solutions for the general two dimensional SDE in (1.1), our main goal is to approximate the dynamics by a suitable one dimensional local version. For this we consider the marginal probability density of the asset, also called the reduced density or effective probability, as our main object of interest. The reduced density Q of F at time t, starting from time t_0, is defined as

$$Q(t,F)\,dF = \mathbb{P}\left[F < \tilde{F}_t < F + dF \,\middle|\, \tilde{F}_{t_0} = f, v_{t_0} = \alpha\right]. \tag{2.1}$$

Given the reduced density for a specified exercise time T, we can then recover the corresponding call and put prices for all strikes by an evaluation of

$$\begin{aligned} C_Q(T,K) &= \int_K^\infty (F-K)Q(T,F)\,dF \qquad \text{and} \\ P_Q(T,K) &= \int_{-\infty}^K (K-F)Q(T,F)\,dF. \end{aligned}$$

To compute the reduced density, we derive a PDE of the form

$$\partial_t Q(t,F) = \partial_{FF}\left[D(t,F)Q(t,F)\right], \quad Q(t_0,f) = \delta(F-f), \tag{2.2}$$

where $D(\cdot,\cdot)$ is a function that involves the model parameters and depends on t and the asset value F. It can be viewed as a local volatility function. To derive this PDE, we use singular perturbation methods to systematically analyze the system

$$\begin{cases} d\tilde{F}_t = \varepsilon C(\tilde{F}_t) v_t\, dW_t^{(1)}, & \tilde{F}_{t_0} = f, \\ dv_t = \mu(v_t)\,dt + \varepsilon\nu(v_t)\,dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)}\rangle_t = \rho\,dt. \end{cases} \tag{2.3}$$

Once the result is derived we set ε back to 1.

The effective PDE, also called the effective forward equation, is accurate to order $\mathcal{O}(\varepsilon^2)$. For achieving a stable and efficient numerical implementation to solve the PDE, we especially need to specify the boundary behaviour. This leads us to consider two PDEs for accumulating probability. The lower boundary b_l is either explicitly specified when the model under consideration is not admissible for values below b_l, or artificially when setting up the grid for numerical computation. The upper boundary is set by specifying the highest level b_u we wish to consider. The corresponding values of the probability densities are denoted by Q^L and Q^R. In our setting the PDEs are given by

$$\begin{aligned} & \partial_t Q^L(t) = \lim_{F \downarrow b_l} \partial_F \left[D(t,F) Q(t,F) \right], && Q^L(t_0) = 0 \\ \text{and} \qquad & \partial_t Q^R(t) = -\lim_{F \uparrow b_u} \partial_F \left[D(t,F) Q(t,F) \right], && Q^R(t_0) = 0. \end{aligned}$$

For the SABR model the derivation can be found in [13, 21], the fSABR model is considered in [20].

Next, we derive the effective equation for the general setting (1.1), then we restrict ourselves to a special class of dynamics, called the ZABR class. This includes the standard ZABR model [1], but also the free boundary version of ZABR (fZABR), which generalizes the results from [4], and the mean-reverting ZABR (mrZABR).

2.1 The Effective PDE

To derive the effective PDE we make the following assumptions related to (1.1):

Assumption I. The drift term, $\mu(\cdot)$, is differentiable, with derivative $\mu'(\cdot)$, and a solution $Y(t, t_0, \alpha)$ to the following PDE exists:

$$\begin{cases} \partial_t Y(t,t_0,\alpha) = \mu(Y(t,t_0,\alpha)) \\ Y(t,t,\alpha) = \alpha \\ Y(t_0,t_0,\alpha) = \alpha. \end{cases}$$

Assumption II. The function Y is differentiable and has an inverse function $y(t_0, t, a)$ such that

$$Y(t,t_0,\alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0,t,a).$$

Remark 2.1. Functions $\mu(\cdot)$ allowing a closed-form solution include:

(i) for $\mu(x) = \mu$ the solution is $Y(t,t_0,\alpha) = \alpha + \mu(t-t_0)$.

(ii) for $\mu(x) = \kappa(\theta - x)$ the solution is $Y(t,t_0,\alpha) = \alpha e^{-\kappa(t-t_0)} + \theta(1 - e^{-\kappa(t-t_0)})$.

Assumption III. The functions

$$X(t,t_0,\alpha)=\partial_\alpha Y(t,t_0,\alpha),\qquad Z(t,u)=Z(t,u,t_0,\alpha)=y(u,t,Y(t,t_0,\alpha)),$$
$$z(F)=\int_f^F\frac{1}{C(u)}\,du,\qquad s(t)=S(t_0,t,\alpha)=\int_{t_0}^t Z(t,u,t_0,\alpha)^2\,du$$

and

$$\psi(t,u,Z)=\nu\big(Z(t,u)\big)Z(t,u)X\big(t,u,Z(t,u)\big)$$

are well defined[1], $X\big(t,u,Z(t,u)\big)^{-1}$ exists, and the following integral functions are defined:

$$I_1(t)=\rho\int_{t_0}^t\psi(t,u,Z)\,du,$$
$$I_2(t)=2\int_{t_0}^t\nu\big(Z(t,u)\big)^2X\big(t,u,Z(t,u)\big)^2\int_u^t Z(t,v)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,$$
$$I_3(t)=\rho\int_{t_0}^t\psi(t,u,Z)\int_u^t Z(t,v)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,$$
$$I_4(t)=\rho^2\int_{t_0}^t\psi(t,u,Z)\int_u^t\partial_Z\Big(\psi(t,v,Z)\Big)X\big(t,v,Z(t,v)\big)^{-1}\,dv\,du,$$
$$I_5(t)=\int_{t_0}^t\nu\big(Z(t,u)\big)^2X\big(t,u,Z(t,u)\big)^2\,du.$$

Assumption IV. The function $C(\cdot)$ is differentiable at f, with derivative denoted by $C'(\cdot)$.

Theorem 2.2. *Given that the general stochastic volatility model (1.1) obeys Assumptions I–IV, an effective PDE for the effective probability (2.1), of the form (2.2), can be derived, with function D given by*

$$D(t,F)=\frac{1}{2}a(t)^2C(F)^2e^{G(t)}\big(1+2b(t)z(F)+c(t)z(F)^2\big),$$

where the coefficients are specified as

$$a(t)=Y(t,t_0,\alpha),\qquad c(t)=b(t)^2+\frac{1}{a(t)s(t)^2}I_2(t)-\frac{6b(t)}{s(t)^2}I_3(t)+\frac{2}{a(t)s(t)^2}I_4(t),$$
$$b(t)=\frac{1}{a(t)s(t)}I_1(t),\quad G(t)=-s(t)c(t)-s(t)b(t)\Gamma_0+\frac{1}{a^2}I_5(t)$$

[1]Note that this definition of $z(F)$ differs by a factor of $\frac{1}{\varepsilon}$ compared to the definition in (A.4).

and

$$\Gamma_0 = -C'(f).$$

Proof. In Appendix A we show that the effective PDE of order $\mathcal{O}(\varepsilon^2)$ is given by

$$\begin{cases} \partial_t Q(t,F) = \dfrac{1}{2}\varepsilon^2 a(t)^2 \partial_{FF}\Big[C(F)^2 Q(t,F) e^{\varepsilon^2 G(t)}\big(1 + 2\varepsilon b(t) z(F) + \varepsilon^2 c(t) z(F)^2\big)\Big] \\ Q(t,F) \to \delta(F-f) \text{ as } t \to t_0^+. \end{cases} \tag{2.4}$$

This comes from combining (A.3), (A.5), (A.10) and (A.1). Setting ε to 1 yields the desired form. □

Remark 2.3. The class of models is not restricted to the above choices of the function Γ_0. Choosing a different form could in turn impose a new version of Assumption IV.

Figure 1 shows the output obtained by numerically solving the effective PDE. It is the density of the asset at maturity and depends on all the input parameters.

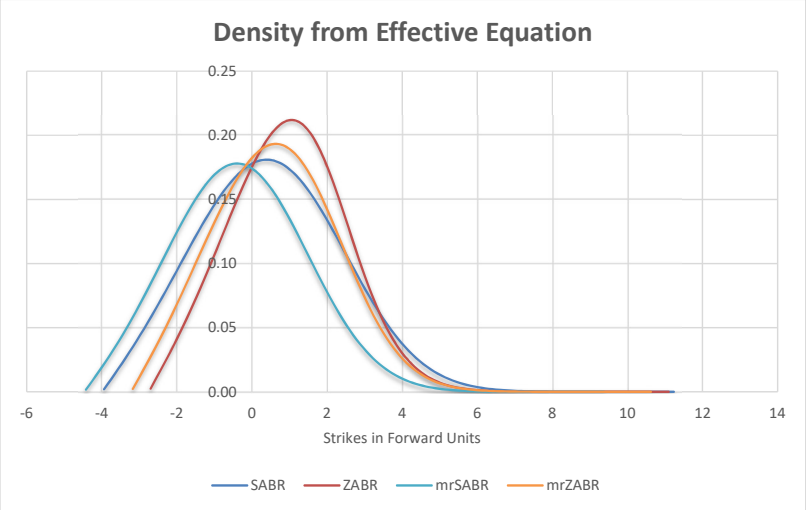


Figure 1: Output from numerically solving the effective PDE for SABR, ZABR, mrSABR and mrZABR.

2.2 Implied Volatility Formula

As seen in Theorem 2.2, the effective PDE of the general model is of the same form as the classical or mean-reverting SABR [10, 16]. In particular, both these models fit into the same framework, which allows for a direct approximation of the implied normal or Black volatility [9].

To show that our model also fits into this framework, we use Effective Media theory [12] to approximate the effective PDE of Equation (2.4) by a suitable SABR effective PDE. To facilitate this we consider a fixed maturity T and set the initial time $t_0 = 0$, from now on. We start by defining time independent parameters $\bar{b}$, $\bar{c}$ and $\bar{G}$, which allow for an approximation of Equation (2.4) to order $\mathcal{O}(\varepsilon^2)$, at time T. These parameters are generally given by [12, Equation 2.4]:

$$\begin{aligned}
\bar{b} &= \frac{2}{T^2}\int_0^T ub(u)\,du \\
\bar{c} &= \frac{3}{T^3}\int_0^T u^2 c(u)\,du + \frac{18}{T^3}\int_0^T b(u)\int_0^u vb(v)\,dv\,du - 3\bar{b}^2 \\
\bar{G} &= \frac{1}{T}\int_0^T G(u)\,du + \frac{1}{T}\int_0^T u(c(u)-\bar{c})\,du.
\end{aligned}$$

Note that before using these equations, the functions b and c of Theorem 2.2 must be modified to fit into the setting of [12, Equation 2.2].

With the constant effective parameters we can define the so called effective SABR parameters:

$$\nu_{\text{eff}} = \sqrt{\bar{c}}, \qquad \rho_{\text{eff}} = \frac{\bar{b}}{\sqrt{\bar{c}}}, \qquad \alpha_{\text{eff}} = \alpha\left(1 + \frac{1}{2}\bar{G} + \frac{1}{4}\alpha\bar{b}\Gamma_0 T\right). \tag{2.5}$$

These allow us to approximate our model to order $\mathcal{O}(\varepsilon^2)$ with a SABR model, and in turn provide us with the various available implied volatility formulae. One prominent example would be the formula provided in [9]:

$$\sigma(T,K) = \varepsilon\frac{\nu_{\text{eff}}(K-f)}{R(\zeta)}\begin{cases} 1+\varepsilon^2\Theta(\zeta)T & \text{if } \Theta \geq 0 \\ \frac{1}{1-\varepsilon^2\Theta(\zeta)T} & \text{if } \Theta < 0, \end{cases}$$

with

$$\begin{aligned}
R(\zeta) &= \log\left(\frac{\rho_{\text{eff}} + \zeta + E(\zeta)}{1+\rho_{\text{eff}}}\right), \\
\Theta(\zeta) &= \frac{\nu_{\text{eff}}^2}{24}\left(3\frac{\rho_{\text{eff}} + \zeta - \rho_{\text{eff}}E(\zeta)}{R(\zeta)E(\zeta)} - 1\right) + \frac{\Delta_0\alpha_{\text{new}}^2}{6}\left(1 - \rho_{\text{eff}}^2 + \frac{(\rho_{\text{eff}}+\zeta)E(\zeta) - \rho_{\text{eff}}}{R(\zeta)}\right),
\end{aligned}$$

where

$$\zeta = \frac{\nu_{\text{eff}}}{\alpha_{\text{new}}} \int_f^K \frac{1}{C(u)}\, du, \qquad E(\zeta) = \sqrt{1 + 2\rho\zeta + \zeta^2},$$
$$\alpha_{\text{new}} = \alpha_{\text{eff}}\big(1 - \frac{1}{4}\rho_{\text{eff}}\nu_{\text{eff}}\alpha_{\text{eff}}\Gamma_0 T\big), \qquad \Delta_0 = \frac{1}{4}C(f)C''(f) - \frac{1}{8}C'(f)^2.$$

In the next section we consider special cases to illustrate the applicability of our approach.

2.3 ZABR-type Models

Given the prevailing negative interest rate regime for some currencies, we focus on stochastic volatility models that are often applied for managing the volatility risk for this asset class. Such models are applied, for instance, to parametrize the implied volatility surface, but also for pricing cash-settled swaptions and constant maturity swaps. Since practitioners have found that standard SABR models are too restrictive for this task, we consider specific choices of the coefficients that lead to ZABR-type models. In particular we consider stochastic volatility models of the form (2.3) where the functions μ, C and ν are of the form given in Table 1.

μ	C	ν	Model
0	$\tilde{F}_t^{\beta}$	v_t	SABR
0	$\tilde{F}_t^{\beta}$	1	Normal SABR (nSABR)
0	$(\tilde{F}_t + d)^{\beta}$	v_t	Displaced SABR
0	$\lvert\tilde{F}_t\rvert^{\beta}$	v_t	Free SABR (fSABR)
$\kappa(\theta - v_t)$	$\tilde{F}_t^{\beta}$	v_t	mean reverting SABR (mrSABR)
0	$\tilde{F}_t^{\beta_1}$	$v_t^{\beta_2}$	ZABR
0	$(\tilde{F}_t + d)^{\beta_1}$	$v_t^{\beta_2}$	Displaced ZABR
0	$\lvert\tilde{F}_t\rvert^{\beta_1}$	$v_t^{\beta_2}$	Free ZABR (fZABR)
$\kappa(\theta - v_t)$	$\tilde{F}_t^{\beta_1}$	$v_t^{\beta_2}$	mean reverting ZABR (mrZABR)

Table 1: Parametrizations of the ZABR-type models.

The SABR-type models were considered in [4, 15, 20, 26] and the ZABR model was introduced in [1]. We consider the displaced ZABR, dZABR, the free ZABR, fZABR, and the mean-reverting ZABR, mrZABR, models. Furthermore, our numerical approach for solving the standard ZABR model differs from the approach in [1] and is arbitrage free by construction.

The models presented in Table 1 are a non-exhaustive list with other choices possible. Different choices for the function C are, for example, suggested in [18] or [19].

Remark 2.4. Note that the absolute value appearing in the function C of the free SABR and free ZABR model is not differentiable at the point 0 and thus would not satisfy the Assumption IV and fit into the framework of Theorem 2.2. To overcome this problem we follow the setting of the original free SABR model [4] and only consider the case where $f \neq 0$.

2.3.1 Effective PDE for ZABR

Consider the ZABR model as specified in Table 1. The corresponding coefficients for the effective PDE in (2.4) may be greatly simplified.

First of all the function Y has an explicit form that equals the initial value:

$$Y(t, t_0, \alpha) = \alpha$$

Evaluation of the coefficients needed for the effective PDE yields

$$\begin{aligned} b &= \rho\nu\alpha^{\beta_2-2}, \\ \tilde{c} &= \nu^2\alpha^{2(\beta_2-2)}(1+(\beta_2-1)\rho^2), \\ G(t) &= -\rho^2\nu^2\alpha^{2(\beta_2-1)}(t-t_0)(\beta_2-1) - (t-t_0)\rho\nu\alpha^{\beta_2}\Gamma_0 \end{aligned} \tag{2.6}$$

and gives the ZABR effective PDE characterized by the function

$$D(t,F) = \frac{1}{2}\alpha^2 C(F)^2(1+2bz(F)+\tilde{c}z(F)^2)e^{G(t)}.$$

For the numerical implementation of the ZABR model, we use the effective SABR parameters presented in Section 2.2. Given the explicit form of the coefficients in (2.6), the effective SABR parameters are given by:

$$\begin{aligned} \nu_{\text{eff}} &= \nu\alpha^{\beta_2-1}\sqrt{1+(\beta_2-1)\rho^2} \\ \rho_{\text{eff}} &= \frac{\rho}{\sqrt{1+(\beta_2-1)\rho^2}} \\ \alpha_{\text{eff}} &= \alpha\left(1+\frac{1}{4}\rho^2\nu^2\alpha^{2(\beta_2-1)}(1-\beta_2)T\right) \end{aligned} \tag{2.7}$$

Note that these are the coefficients used for the actual implementation. In particular, we have already set $\varepsilon = 1$ and approximated the original exponential

function. To guarantee that the term $\sqrt{D(t,F)}$ remains real, we further impose the condition

$$\beta_2 > 1 + \frac{\rho^2 - 1}{\rho^2}.$$

Since we are using a SABR model to approximate the ZABR model, it is not surprising that we observe similar behaviour of the models. For example, if we shift the underlying forward rate f, as done in Figure 2, we can see that the implied volatility moves in same direction as the forward.

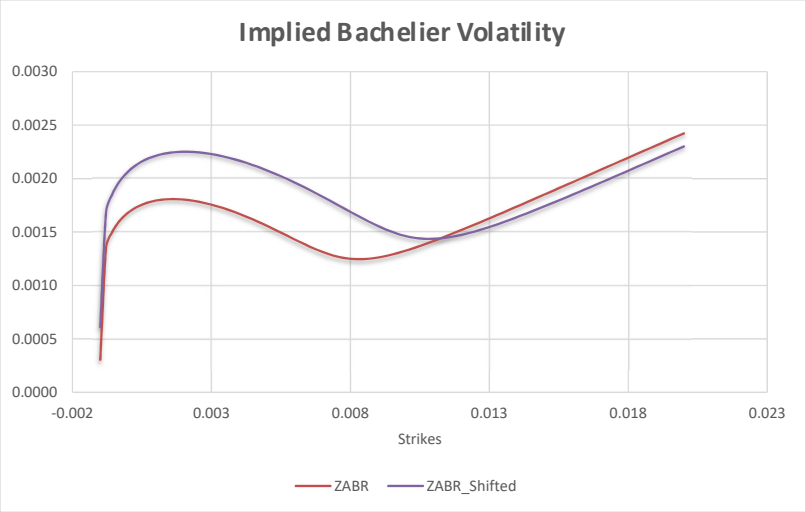


Figure 2: The implied volatility for the ZABR model with parameters $\beta = 0.5$, $\beta_2 = 0.8$, $\nu = 0.3$, $\rho = -0.8$, an underlying forward rate of $f = 0.005$, which is shifted by 0.002, and a displacement of 0.001.

2.3.2 Effective PDE for fZABR and mrZABR

Having examined the ZABR model, we can now consider the natural extensions of the free ZABR and mean-reverting ZABR models.

For the free ZABR model, the effective SABR parameters are the same as for the ZABR model and are given in (2.7). The difference between the models lies exclusively in the function $C(F)$, as was the case in the extension of the SABR to the free SABR model, see [20].

Considering the mean-reverting ZABR model with a reversion back to the initial state, i.e., with $\theta = \alpha$, as specified in Table 1, the corresponding parameters for

the effective forward equation change and are given by

$$
\begin{aligned}
b(t) &= \frac{\rho\nu\alpha^{\beta_2-2}}{\kappa(t-t_0)}\big(1-e^{-\kappa(t-t_0)}\big),\\
c(t) &= \frac{(1+\rho^2)\nu^2\alpha^{2(\beta_2-2)}}{\kappa^2(t-t_0)^2}\big(1-e^{-\kappa(t-t_0)}\big)^2\\
&\quad+\frac{6\rho^2\nu^2\alpha^{2(\beta_2-2)}}{\kappa^3(t-t_0)^3}\big(1-e^{-\kappa(t-t_0)}\big)\big(1-\kappa(t-t_0)-e^{-\kappa(t-t_0)}\big)\\
&\quad+(1+\beta_2)\frac{2\rho^2\nu^2\alpha^{2(\beta_2-2)}}{\kappa^2(t-t_0)^2}\big(1-(1+\kappa(t-t_0))e^{-\kappa(t-t_0)}\big),\\
G(t) &= -\alpha^2(t-t_0)c-\frac{\rho\nu\alpha^{\beta_2}}{\kappa}\big(1-e^{-\kappa(t-t_0)}\big)\Gamma_0+\frac{\nu^2\alpha^{2(\beta_2-1)}}{2\kappa}\big(1-e^{-2\kappa(t-t_0)}\big).
\end{aligned}
$$

Fixing a specified maturity T, the corresponding constant effective parameters of Section 2.2 are given by

$$
\begin{aligned}
\bar{b} &= \frac{2\rho\nu\alpha^{\beta_2-1}}{\kappa^2T^2}\big(\kappa T-1+e^{-\kappa T}\big),\\
\bar{c} &= \frac{3(1+\rho^2)\nu^2\alpha^{2(\beta_2-1)}}{2(\kappa T)^3}\big(2\kappa T+4e^{-\kappa T}-3-e^{-2\kappa T}\big)\\
&\quad+6(1+\beta_2)\frac{\rho^2\nu^2\alpha^{2(\beta_2-1)}}{(\kappa T)^3}\big(\kappa T+2e^{-\kappa T}-2+\kappa Te^{-\kappa T}\big)\\
&\quad-12\rho^2\nu^2\alpha^{2(\beta_2-1)}\left(\frac{\kappa T-1+e^{-\kappa T}}{(\kappa T)^2}\right)\\
\bar{G} &= \frac{\nu^2\alpha^{2(\beta_2-1)}}{4\kappa^2T}\big(2\kappa T+e^{-2\kappa T}-1\big)-\frac{1}{2}\bar{c}T-\frac{\rho\nu\alpha^{\beta_2}}{\kappa^2T}\big(\kappa T-1+e^{-\kappa T}\big)\Gamma_0.
\end{aligned}
$$

Here we can now use (2.5) to derive the effective SABR parameters. As seen in Figure 3 and 4, in both cases we still observe that the implied volatility moves together with the shifted underlying.

3 Numerical Examples

For our numerical implementation we use the effective SABR parameters derived for each model and the methods described in [13, 20, 21] adapted to our setting.

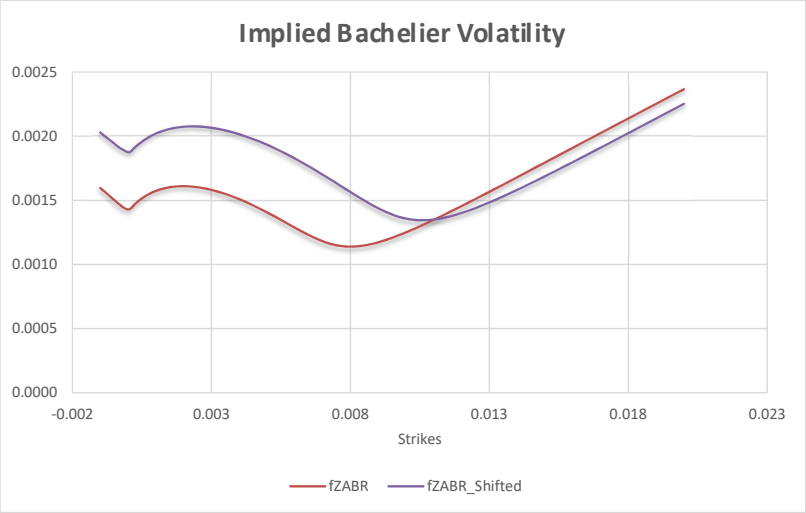


Figure 3: Free ZABR implied volatility with parameters as in Figure 2.

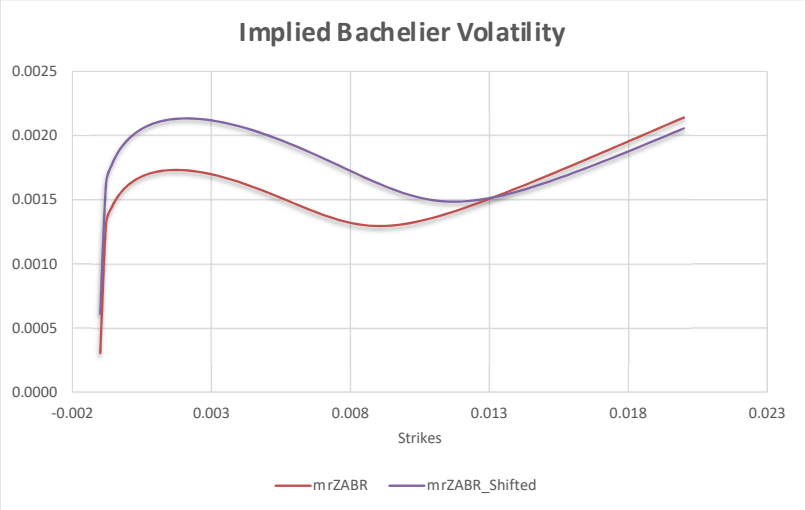


Figure 4: Mean reverting ZABR implied volatility with parameters as in Figure 2 and $\kappa = 0.2$.

3.1 Comparison SABR and ZABR Type Models

To see the impact of our new models, we compare the implied volatilities generated by each model. For this we use the same parameters as in Figure 2 and various values for the new parameters β_2 and κ. The results are shown in Figure 5 and Figure 6. As we can see, the change from a SABR to a ZABR model heavily

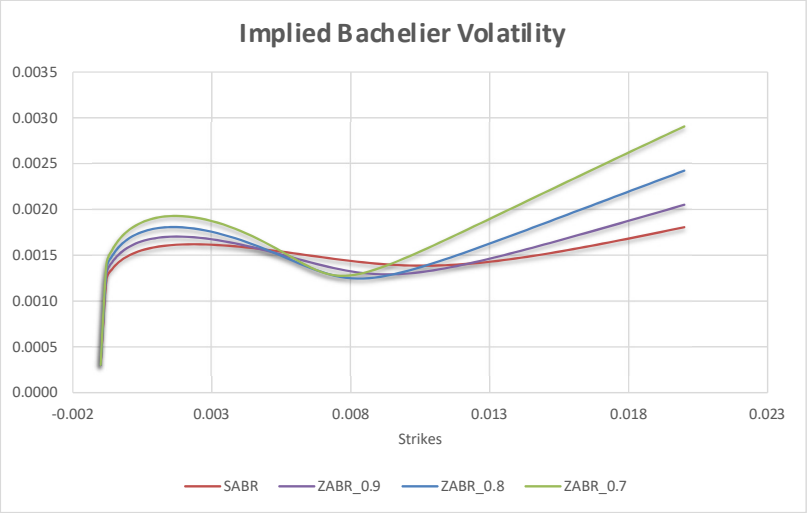


Figure 5: Implied volatility for the ZABR model when β_2 changes.

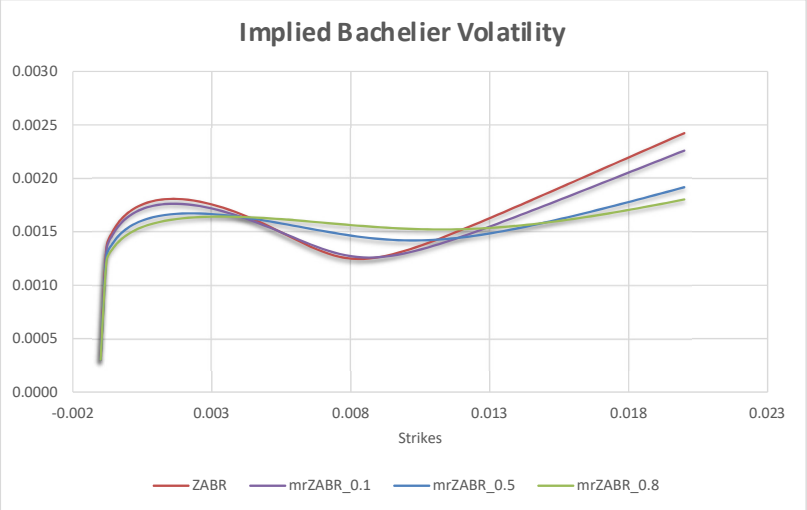


Figure 6: Implied volatility for the mean reversion ZABR when κ changes.

influences the smile. In particular the OTM end shows much steeper behaviour. The inclusion of mean reversion again works in the contrary direction and brings the model closer to the SABR case.

In Figure 7 we considered the density of the ZABR model. Here we can see that the model exhibits a much higher and steeper peak in the density function, compared with the SABR model.

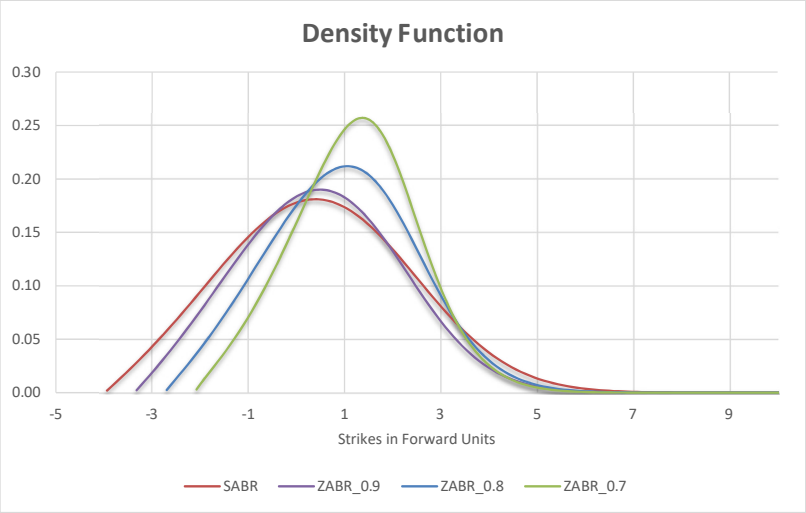


Figure 7: Density for the ZABR model when β_2 changes.

Finally, Figure 8 shows that the implied volatility formulas yield a good approximation to the models for moderate to high strikes. As in the case of the SABR model, the approximation becomes worse for low strikes.

3.2 Calibration

One crucial aspect of a model is its ability to fit to real market data. One common way to calibrate a model is to consider a set of implied volatilities and try to adjust the model parameters to minimize the error between market and model volatilities. For our numerical experiment we calibrated the models on a set of normal swaption volatilities with maturity and tenor of 5 years. The volatility data is for the EUR currency and corresponds to the dates 2 Sept 2019 and 1 Nov 2019.

Since we are approximating the various models with a version of the SABR model, it is assumed that the fit is of the same quality. In fact, if we consider the ZABR model, we can predetermine the new parameter β_2 based on a market analysis and calibrate the remaining parameters to swaption data. In Figures 9 and 10 we can see that the quality of the fit matches the SABR model for various dates and fixed β_2. Also, for the mean-reverting ZABR, in Figure 9, we observe a calibration result similar to that obtained for the SABR model. Overall a good fit of the models to the market data is observed.

If the quality of the calibration is similar to that of the SABR model, indeed

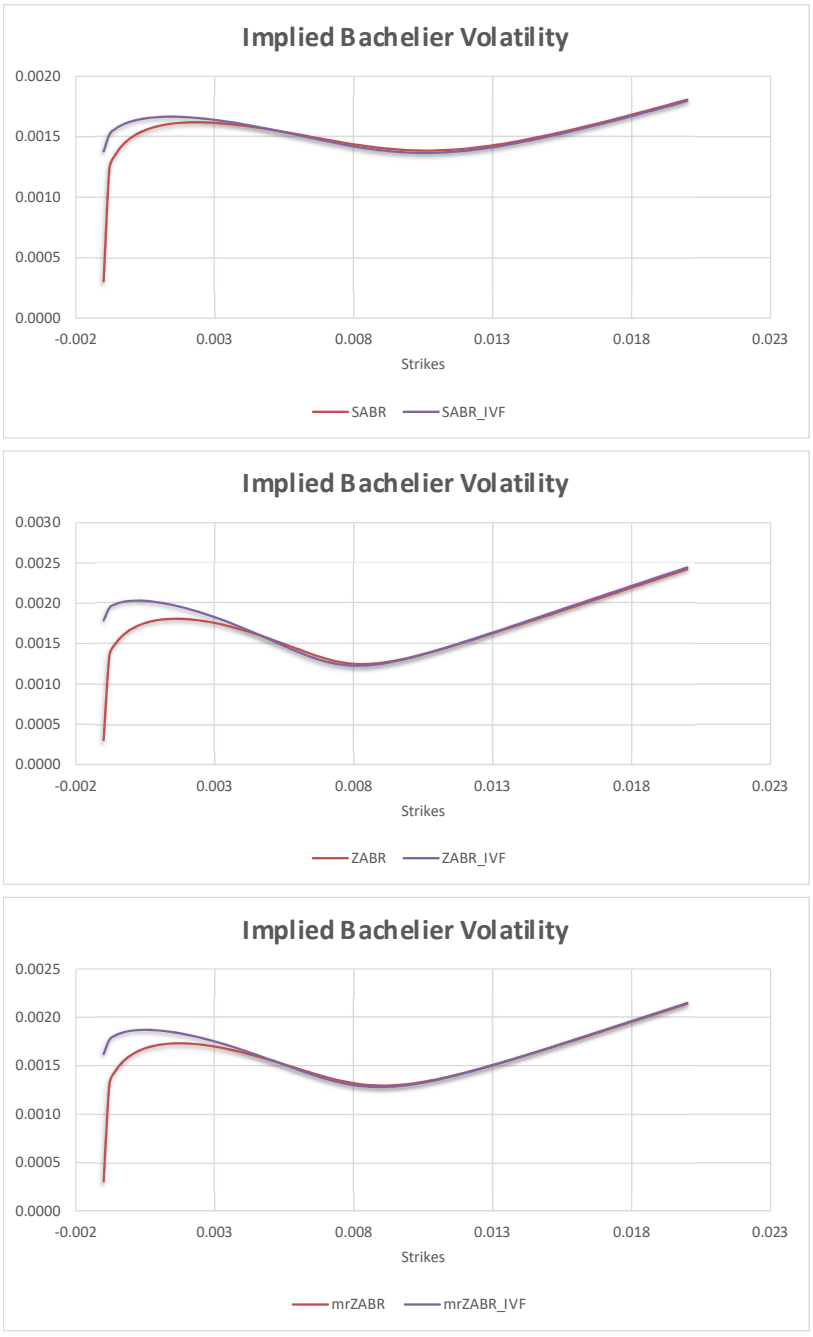
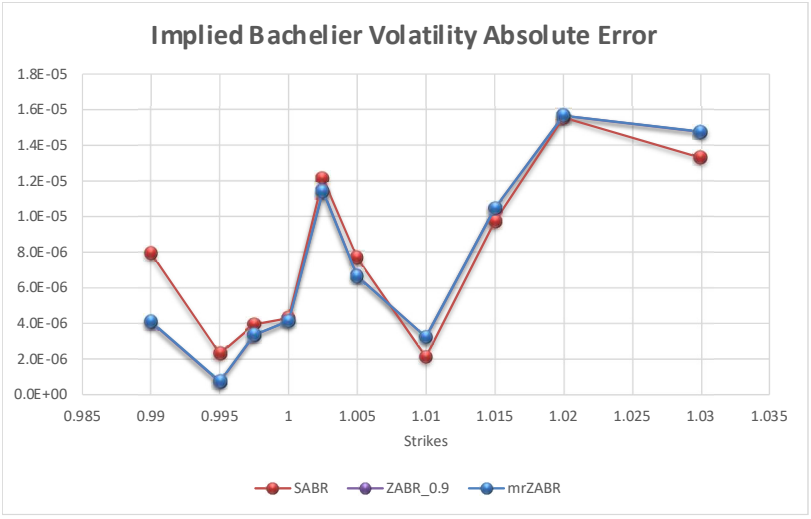


Figure 8: Comparison of implied Bachelier volatility functions for the SABR, ZABR and mrZABR models.



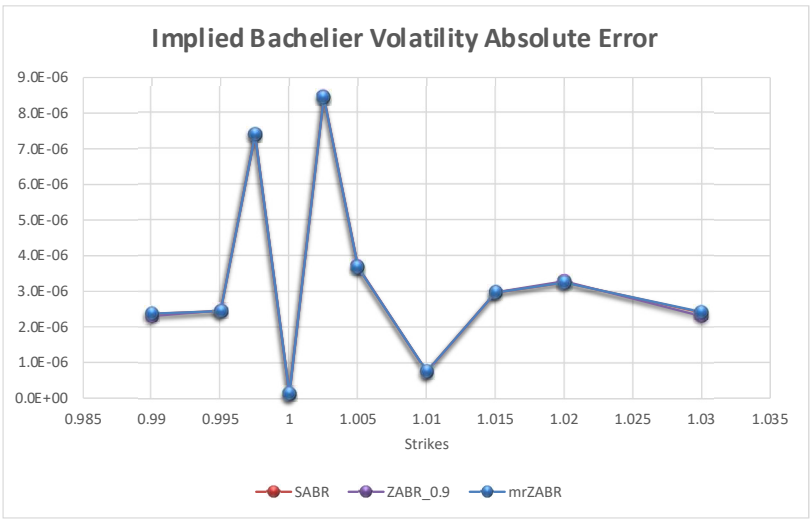


Figure 9: The ZABR model with parameter $\beta_2 = 0.9$ and the mrZABR model with $\kappa = 0.3$ calibrated to two different dates.

we may ask ourselves why we are not using the SABR to begin with. Here it is relevant to mention that in most applications, we are not only interested in matching the current market data, but we also want to provide a set of economic scenarios for future prices or obligations that may be more complicated and even path dependent. In [5] it is shown that models with the same calibration quality may reveal their flexibility when exotic derivatives are considered. For instance, CMS index related derivatives illustrate this. We do not repeat this study here.

Using, for example, an Euler-Maruyama scheme to simulate the model, the ad-

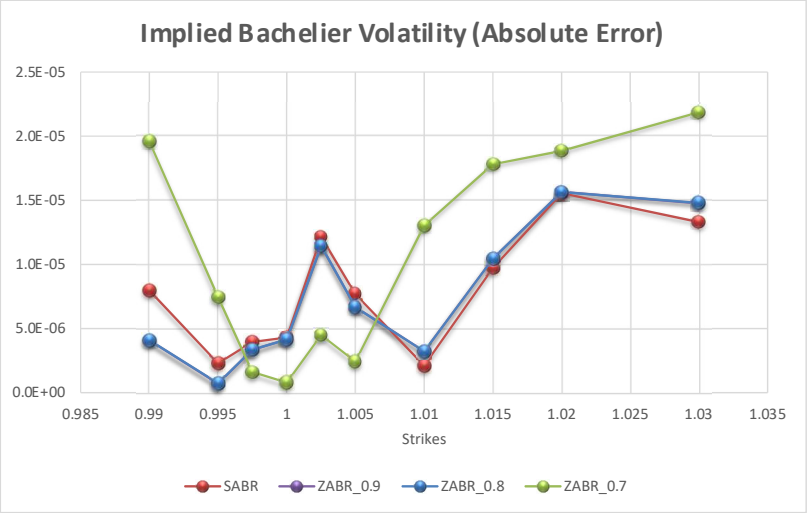


Figure 10: The ZABR model with different values of β_2 calibrated to one date.

ditional parameters play an important role, since, at each time step, we would need a different SABR model to approximate the current distribution. We shall further investigate the additional benefits coming from such considerations in a future study.

3.3 Comparison to MC

For the comparison with a Monte Carlo experiment we chose the same parameter set as above and ran a simulation with one million paths with 240 time steps for all the models except for the free SABR and ZABR models where we used ten million paths with 480 time steps. Figure 11 shows the difference in implied Bachelier volatility observed. The results from using both approaches are very close. Here we observe an average relative error of about 5%, with the most significant contribution in the tails.

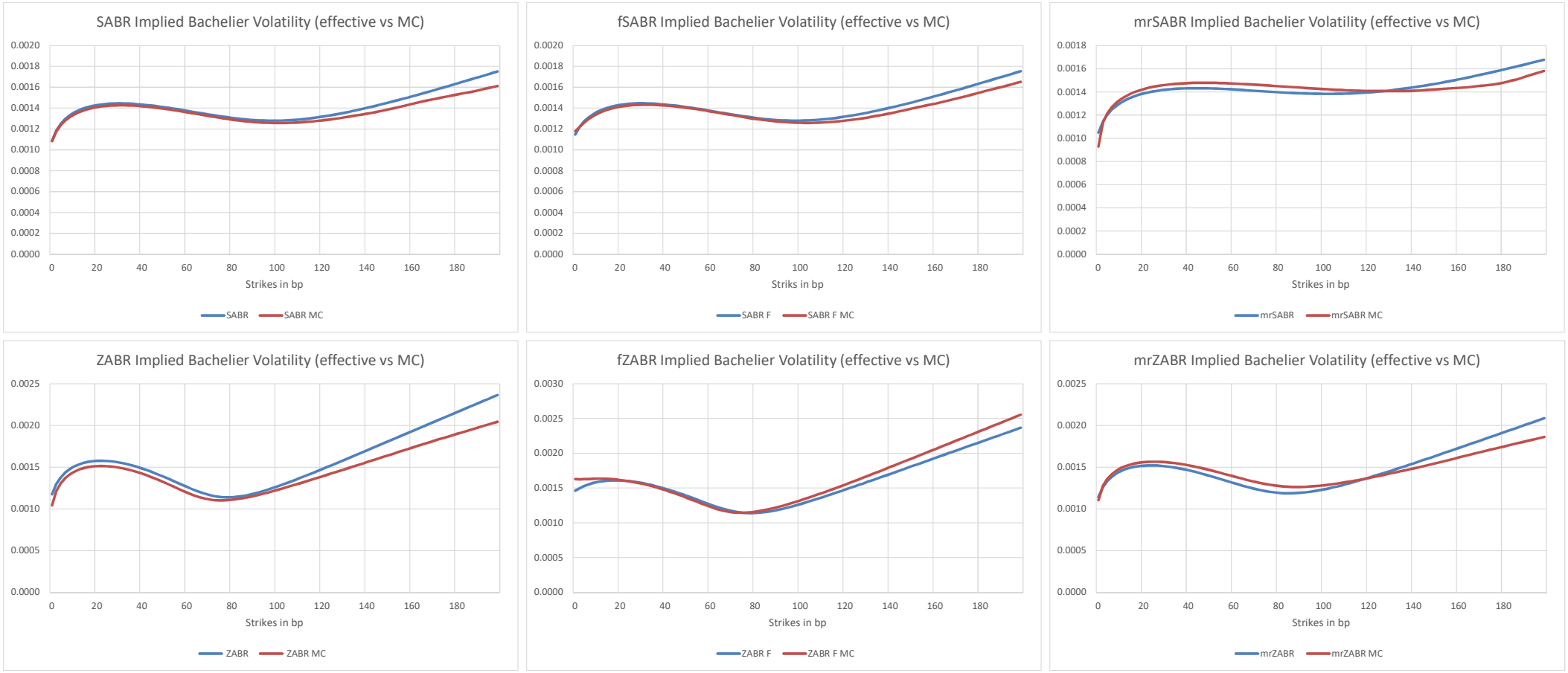


Figure 11: Implied Bachelier volatility computed from the Call option prices obtained from the effective equation and Monte Carlo simulation for the SABR (top left), ZABR (top right), mrSABR (mid left), mrZABR (mid right), fSABR (bottom left) and fZABR (bottom right).

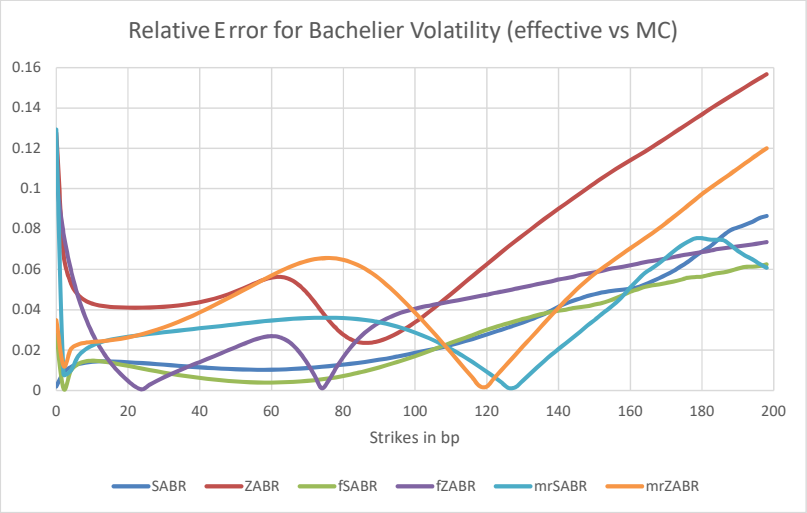


Figure 12: Relative error for the implied Bachelier volatility computed from the call option prices obtained from the effective equation and Monte Carlo simulation for the SABR, ZABR, mrSABR, mrZABR, fSABR and fZABR.

Figure 12 shows the relative error as a function of the strike and for all the models considered.

4 Conclusion and Summary

We have outlined an approach applicable to a large class of stochastic volatility models. The approach is based on an effective PDE associated with the stochastic volatility model. In particular we have derived a (one dimensional) local volatility representation of the (two dimensional) model. Then, we applied accurate and efficient numerical schemes to calculate option prices. General conditions for applicability were derived. We extend the modeling to include the ZABR model and introduced the free ZABR and mean-reverting ZABR models. The flexibility of the new models was explored and a numerical comparison with existing methods was given.

To derive the results we applied singular perturbation theory. There are, however, other techniques that may be used to derive a local volatility model associated with a stochastic volatility model. For instance we still wish to compare our results to those derived via the method of Markovian Projection, see e.g. [2, 3, 28]. From a numerical methods perspective, we applied the method considered in [13, 20, 21]. While we have included a basic comparison to Monte Carlo simulation results we

still wish to derive numerical schemes for the two dimensional equation as in [32]. The generalization of the numerical scheme derived in the context of analyzing the Heston model seems to be a good starting point for further analyzing the quality of our approximation using effective equations.

A Appendix: Derivations

A.1 Deriving the Effective Forward Equation

As described in Section 2, we analyze a stochastic volatility model of the form

$$\begin{cases} d\tilde{F}_t = \varepsilon C(\tilde{F}_t) v_t \, dW_t^{(1)}, & \tilde{F}_{t_0} = f, \\ dv_t = \mu(v_t)\, dt + \varepsilon \nu(v_t)\, dW_t^{(2)}, & v_{t_0} = \alpha, \\ \text{with} \quad d\langle W^{(1)}, W^{(2)} \rangle_t = \rho \, dt. \end{cases}$$

Following Hagan et al. [8,10,16], we define the probability density $p(t_0, f, \alpha, t, F, A)$ that $\tilde{F}(t) = F$ and $v_t = A$ at time t, given that $\tilde{F}(t_0) = f$ and $v_{t_0} = \alpha$ at time t_0. Furthermore, we define the moments

$$Q^{(k)}(t_0, f, \alpha, t, F) = \int_0^\infty A^k p(t_0, f, \alpha, t, F, A)\, dA$$

and set

$$Q(t, F) = Q^{(0)}(t_0, f, \alpha, t, F) \tag{A.1}$$

to be the reduced probability density of the model given t_0, f and α.

A.2 Volatility Drift

Before analyzing the corresponding PDE equations, let us first take a look at the drift term of the volatility function. For this, we consider the following PDE given by the drift term

$$\begin{cases} \partial_t Y(t, t_0, \alpha) = \mu(Y(t, t_0, \alpha)) \\ Y(t, t, \alpha) = \alpha \\ Y(t_0, t_0, \alpha) = \alpha. \end{cases} \tag{A.2}$$

By Assumptions I and II we know that there exists a solution $Y(t, t_0, \alpha)$ and an inverse function $y(t_0, t, a)$, such that

$$Y(t, t_0, \alpha) = a \qquad \Leftrightarrow \qquad \alpha = y(t_0, t, a).$$

Considering the integrated form of the PDE (A.2), the derivatives of $Y(t, t_0, \alpha)$ satisfy

$$\partial_{t_0} Y(t, t_0, \alpha) = -\mu(\alpha) + \int_{t_0}^{t} \mu'(Y(s, t_0, \alpha)) \partial_{t_0} Y(s, t_0, \alpha)\, ds$$

$$\text{and} \qquad \partial_{\alpha} Y(t, t_0, \alpha) = 1 + \int_{t_0}^{t} \mu'(Y(s, t_0, \alpha)) \partial_{\alpha} Y(s, t_0, \alpha)\, ds.$$

This, in turn, implies that

$$\partial_{t_0} Y(t, t_0, \alpha) = -\mu(\alpha) \partial_{\alpha} Y(t, t_0, \alpha).$$

Remark A.1. Note that in both cases of the Remark 2.1 we have that Y coincides with the expected volatility:

$$Y(t, t_0, \alpha) = \mathbb{E}\big[v_t \,|\, v_{t_0} = \alpha\big].$$

If μ is a non-linear function, this would not be true anymore, since we would have to exchange the expectation with the function μ, which in general can not be done.

A.3 The Forward Equation

Now, we start by considering the Kolmogorov forward equation to get

$$\begin{aligned}\partial_t p(t, F, A) = &-\partial_A \Big[\mu(A) p(t, F, A)\Big] + \frac{1}{2}\varepsilon^2 \partial_{FF}\Big[C(F)^2 A^2 p(t, F, A)\Big] \\ &+\varepsilon^2 \rho \partial_{FA}\Big[C(F) A \nu(A) p(t, F, A)\Big] + \frac{1}{2}\varepsilon^2 \partial_{AA}\Big[\nu(A)^2 p(t, F, A)\Big],\end{aligned}$$

where we have abbreviated $p(t_0, f, \alpha, t, F, A)$ as $p(t, F, A)$. Integrating over A and considering reflecting boundary conditions to conserve the probability, as done for example in [10], we get

$$\begin{aligned}\int_0^{\infty} \partial_A \Big[\mu(A) p(t, F, A)\Big] dA &= \Big[\mu(A) p(t, F, A)\Big]\Big|_0^{\infty} = 0 \\ \int_0^{\infty} \partial_{FA}\Big[C(F) A \nu(A) p(t, F, A)\Big] dA &= \partial_F \Big[C(F) A \nu(A) p(t, F, A)\Big]\Big|_0^{\infty} = 0 \\ \int_0^{\infty} \partial_{AA}\Big[\nu(A)^2 p(t, F, A)\Big] dA &= \partial_A \Big[\nu(A)^2 p(t, F, A)\Big]\Big|_0^{\infty} = 0.\end{aligned}$$

With this, we get the forward equation

$$\begin{cases} \partial_t Q^{(0)}(t, F) = \dfrac{1}{2}\varepsilon^2 \partial_{FF}\Big[C(F)^2 Q^{(2)}(t, F)\Big] \\ Q^{(0)}(t, F) \to \delta(F - f) \text{ as } t \to t_0^+, \end{cases} \tag{A.3}$$

for $t > t_0$

A.4 The Backward Equation

Next, let us consider the Kolmogorov backward equation for $Q^{(k)}$ given by

$$\begin{cases} -\partial_{t_0} Q^{(k)} = \mu(\alpha)\partial_\alpha Q^{(k)} + \frac{1}{2}\varepsilon^2\alpha^2\Big[C(f)^2\partial_{ff}Q^{(k)} + 2\rho\frac{\nu(\alpha)}{\alpha}C(f)\partial_{f\alpha}Q^{(k)} \\ \qquad\qquad + \left(\frac{\nu(\alpha)}{\alpha}\right)^2\partial_{\alpha\alpha}Q^{(k)}\Big] \\ Q^{(k)} \to \alpha^k\delta(F-f) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $Q^{(k)}(t_0, f, \alpha, t, F)$ as $Q^{(k)}$. To cancel the drift term, let us change variables from α to $a = Y(t, t_0, \alpha)$. As seen in Section A.2 the change of variables is provided by

$$\begin{aligned} &\partial_\alpha \to \partial_\alpha Y(t, t_0, \alpha)\partial_a = X(t, t_0, y(t_0, t, a))\partial_a, \\ &\partial_{t_0} \to \partial_{t_0} - \mu(\alpha)\partial_\alpha Y(t, t_0, \alpha)\partial_a = \partial_{t_0} - \mu(y(t_0, t, a))X(t, t_0, y(t_0, t, a))\partial_a. \end{aligned}$$

Here, we use the function X as an abbreviation of $\partial_\alpha Y(t, t_0, \alpha)$. With this the drift term vanishes and we get

$$\begin{cases} -\partial_{t_0} Q^{(k)} = \frac{1}{2}\varepsilon^2 y(t_0, t, a)^2\Big[C(f)^2\partial_{ff}Q^{(k)} + 2\rho\tilde{\nu}C(f)\partial_{f\alpha}Q^{(k)} + \tilde{\nu}^2\partial_{\alpha\alpha}Q^{(k)}\Big] \\ Q^{(k)}(t_0, f, a) \to a^k\delta(F-f) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $Q^{(k)}(t_0, f, a, t, F)$ as $Q^{(k)}$, and

$$\tilde{\nu} = \tilde{\nu}(t_0, t, a) = \frac{\nu(y(t_0, t, a))}{y(t_0, t, a)}X(t, t_0, y(t_0, t, a)).$$

Note that this equation corresponds to the one in [16, A.19]. The only difference is the form of $\tilde{\nu}(t_0, t, a)$ where we have an additional dependence on $y(t_0, t, a)$. Thus, until we reach the point where the explicit form of $\tilde{\nu}$ is needed, our reasoning is the same that used by Hagan et al. [8, 16]. Nevertheless, for completeness, we briefly present the main steps.

We start by changing the variable f to

$$z = \frac{1}{\varepsilon}\int_f^F \frac{1}{C(u)}\,du \tag{A.4}$$

and re-scale $Q^{(k)}$ to be

$$Q^{(k)}(t_0, f, a, t, F) = \frac{a^k}{\varepsilon B(0)}\tilde{Q}^{(k)}(t_0, z, a, t, F) \tag{A.5}$$

with variables

$$B(\varepsilon z) = C(f) \qquad \text{and} \qquad \Gamma(\varepsilon z) = \frac{B'(\varepsilon z)}{B(\varepsilon z)}.$$

This yields the equation

$$\begin{cases} -\partial_{t_0}\tilde{Q}^{(k)} = y(t_0,t,a)^2\Big[\frac{1}{2}\partial_{zz}\tilde{Q}^{(k)} - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z\tilde{Q}^{(k)} - \varepsilon k\rho\frac{\tilde{\nu}}{a}\partial_z\tilde{Q}^{(k)} \\ \qquad\qquad + \frac{1}{2}\varepsilon^2 k(k-1)\frac{\tilde{\nu}^2}{a^2}\tilde{Q}^{(k)} - \varepsilon\rho\tilde{\nu}\partial_{za}\tilde{Q}^{(k)} \\ \qquad\qquad + \varepsilon^2 k\frac{\tilde{\nu}^2}{a}\partial_a\tilde{Q}^{(k)} + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\partial_{aa}\tilde{Q}^{(k)}\Big] \\ \tilde{Q}^{(k)} \to \delta(z) \text{ as } t_0 \to t^-, \end{cases}$$

where we have abbreviated $\tilde{Q}^{(k)}(t_0, z, a, t, F)$ as $\tilde{Q}^{(k)}$.

Now, to cancel the a dependent term $y(t_0,t,a)^2$ in front of every term, we change the time scaling from t_0 to s given by

$$s = S(t_0,t,a) = \int_{t_0}^{t} y(u,t,a)^2\,du.$$

Again, we consider the inverse denoted by $\tilde{t_0}$ such that $t_0 = \tilde{t_0}(s,t,a)$ when $s = S(t_0,t,a)$. Then we get the new PDE

$$\begin{cases} \partial_s\tilde{Q}^{(k)} = \frac{1}{2}\partial_{zz}\tilde{Q}^{(k)} - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z\tilde{Q}^{(k)} - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\,\partial_{zs}\tilde{Q}^{(k)} + \partial_{za}\tilde{Q}^{(k)}\Big) \\ \qquad + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\,\partial_{ss}\tilde{Q}^{(k)} + \partial_{aa}s\,\partial_s\tilde{Q}^{(k)}\Big) - \varepsilon k\rho\frac{\tilde{\nu}}{a}\partial_z\tilde{Q}^{(k)} \\ \qquad + \varepsilon^2 k\frac{\tilde{\nu}^2}{a}\partial_a s\,\partial_s\tilde{Q}^{(k)} + \frac{1}{2}\varepsilon^2 k(k-1)\frac{\tilde{\nu}^2}{a^2}\tilde{Q}^{(k)} \\ \qquad + \varepsilon^2 k\frac{\tilde{\nu}^2}{a}\partial_a\tilde{Q}^{(k)} + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(2\partial_a s\,\partial_{sa}\tilde{Q}^{(k)} + \partial_{aa}\tilde{Q}^{(k)}\Big) \\ \tilde{Q}^{(k)} \to \delta(z) \text{ as } s \to 0^+, \end{cases} \tag{A.6}$$

where we have abbreviated $\tilde{Q}^{(k)}(s, z, a, t, F)$ as $\tilde{Q}^{(k)}$. Note that this equation corresponds to the ones in for example [16, A.29] or [8, A.28]. Now, we argue that to leading order we only have the heat equation

$$\begin{cases} \partial_s\tilde{Q}^{(k)} = \frac{1}{2}\partial_{zz}\tilde{Q}^{(k)} \\ \tilde{Q}^{(k)} \to \delta(z) \text{ as } s \to 0^+. \end{cases} \tag{A.7}$$

Thus, if we expand $\tilde{Q}^{(k)} = \tilde{Q}_0^{(k)} + \varepsilon\tilde{Q}_1^{(k)} + \varepsilon^2\tilde{Q}_2^{(k)} + \cdots$ we can conclude that the first term $\tilde{Q}_0^{(k)}$ is given by the solution of (A.7) as

$$\tilde{Q}_0^{(k)}(s,z,a) = \tilde{Q}_0^{(k)}(s,z) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{z^2}{2s}}.$$

In particular, $\tilde{Q}_0^{(k)}$ does not depend on a and thus $\partial_a \tilde{Q}^{(k)}$ is actually of order $\mathcal{O}(\varepsilon)$ and the last two terms of (A.6) are of order $\mathcal{O}(\varepsilon^3)$ and can be neglected.

A.5 Computing $\tilde{Q}^{(2)}$

With (A.6) we can set up the two PDEs for $\tilde{Q}^{(0)}$ and $\tilde{Q}^{(2)}$. Without explicitly writing them down, we can still see that for $k = 2$ the PDE contains all the terms of $k = 0$ and the additional terms of the last lines. Since our goal is to express $\tilde{Q}^{(2)}$ in terms of $\tilde{Q}^{(0)}$ and to insert this back into the forward equation, we use the ansatz

$$\tilde{Q}^{(2)}(s,z,a,t,F) = H(s,z,a,t,F) e^{2\varepsilon b(s,a)z + \varepsilon^2 \tilde{c}(s,a)z^2 + \varepsilon^2 G(s,\alpha)}$$

and determine the coefficients in such a way that the PDE for H corresponds to the one of $\tilde{Q}^{(0)}$.

This yields the PDE given by

$$\begin{aligned}
\partial_s H = {} & \frac{1}{2}\partial_{zz} H - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z H - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\, \partial_{zs} H + \partial_{za} H\Big) \\
& + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\, \partial_{ss} H + \partial_{aa} s\, \partial_s H\Big) - \left(s\partial_s b + b - \frac{\rho\tilde{\nu}}{a}\right)\frac{2\varepsilon z}{s} H \\
& - \Big(s\partial_s \tilde{c} + 2\tilde{c} - 2\rho\tilde{\nu}\Big(\partial_a b + \partial_s b\, \partial_a s\Big)\Big)\frac{\varepsilon^2 z^2}{s} H \\
& + \left(- \partial_s G + \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}}{a} b + \frac{\tilde{\nu}^2}{a^2} - b\Gamma_0 - 2\rho\tilde{\nu}\Big(\partial_a b + \partial_s b\, \partial_a s\Big)\right)\varepsilon^2 H \\
& + 2\left(b - \frac{\rho\tilde{\nu}}{a}\right)\varepsilon\left(\partial_z H + \frac{z}{s} H\right) - 2\left(\rho\tilde{\nu} b - \frac{\tilde{\nu}^2}{a}\right)\varepsilon^2 \partial_a s\, \partial_s H,
\end{aligned} \tag{A.8}$$

where we have abbreviated $H(s,z,a,t,F)$ as H. Here $\Gamma_0 = \Gamma(0)$ is a suitable approximation of $\Gamma(\varepsilon z)$ with order $\mathcal{O}(\varepsilon^2)$. To bring the equation of H closer to the one for $\tilde{Q}^{(0)}$ we set b such that

$$s\partial_s b + b - \frac{\rho\tilde{\nu}}{a} = 0,$$

in which case b has the form

$$b(s,a) = \frac{1}{as} I_1(s,t,a)$$
$$\text{with} \qquad I_1 = I_1(s,t,a) = \rho \int_0^s \tilde{\nu}\big(\tilde{t_0}(x,t,a),t,a\big)\, dx.$$

Switching the integration variable from x to $u = \tilde{t_0}(x,t,a)$, $I_1(s,t,a)$ may be written as

$$\begin{aligned} I_1(s,t,a) &= \rho \int_{\tilde{t_0}(s,t,a)}^t \tilde{\nu}\big(u,t,a\big) y^2(u,t,a)\, du \\ &= \rho \int_{\tilde{t_0}(s,t,a)}^t \nu\big(y(u,t,a)\big) y(u,t,a) X(t,u,y(u,t,a))\, du. \end{aligned}$$

Note that here the only s dependence of the integral I_1 comes from the boundary $\tilde{t_0}(s,t,a)$. To handle the term containing $\varepsilon(\partial_z H + \frac{z}{s}H)$ we recall that

$$H_0(s,z) = \frac{1}{\sqrt{2\pi s}} e^{-\frac{z^2}{2s}}$$

and that $\partial_z H^0(s,z) + \frac{z}{s} H^0(s,z) = 0$. With this the PDE of order $\mathcal{O}(\varepsilon)$ is

$$\partial_s H = \frac{1}{2}\partial_{zz} H - \frac{1}{2}\varepsilon \Gamma_0 \partial_z H - \varepsilon \rho \tilde{\nu} \partial_a s\, \partial_{zs} H.$$

As in [16] we can show through the concrete form of H^0 that H^1 has the form

$$H^1(s,z) = -\frac{1}{2} s \Gamma_0 \partial_z H^0 - I_3(s,t,a) \partial_{zzz} H^0$$
$$\text{with} \qquad I_3(s,t,a) = \frac{1}{2}\rho \int_0^s \tilde{\nu}(\tilde{t_0}(x,t,a),t,a) \partial_a S(\tilde{t_0}(x,t,a),t,a)\, dx.$$

Using our new form of $\tilde{\nu}$ and the explicit form of $\partial_a S(\tilde{t_0}(x,t,a),t,a)$ given by

$$\begin{aligned} \partial_a S(\tilde{t_0},t,a) &= 2\int_{\tilde{t_0}}^t y(u,t,a) \partial_a y(u,t,a)\, du \\ &= 2\int_{\tilde{t_0}}^t y(u,t,a) X(t,u,y(u,t,a))^{-1}\, du, \end{aligned}$$

we get

$$\begin{aligned} I_3 = I_3(s,t,a) = \rho \int_{\tilde{t_0}(s,t,a)}^t & \nu(y(u,t,a)) y(u,t,a) X(t,u,y(u,t,a)) \\ & \times \int_u^t y(v,t,a) X(t,v,y(v,t,a))^{-1}\, dv\, du. \end{aligned}$$

Moreover, we can conclude as in [16] that

$$\varepsilon\left(\partial_z H + \frac{z}{s}H\right) = \frac{3I_3}{s^2}\frac{\varepsilon^2 z^2}{s}H - \left(\frac{3I_3}{s^2} - \frac{1}{2}\Gamma_0\right)\varepsilon^2 H + \mathcal{O}(\varepsilon^3).$$

Before determining the parameters $\tilde{c}$ and G in (A.8) we rewrite the last term in terms of H. For this note that to order $\mathcal{O}(\varepsilon^2)$ we have

$$\varepsilon^2\partial_s H = \frac{1}{2}\varepsilon^2 s\partial_{zz}H = \frac{1}{2}\varepsilon^2\left(\frac{z^2}{s^2} - \frac{1}{s}\right)H.$$

With this, the PDE in (A.8) simplifies to

$$\begin{aligned}
\partial_s H = {} & \frac{1}{2}\partial_{zz}H - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z H - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\,\partial_{zs}H + \partial_{za}H\Big) \\
& + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\,\partial_{ss}H + \partial_{aa}s\,\partial_s H\Big) \\
& - \Bigg[s\partial_s\tilde{c} + 2\tilde{c} - 2\rho\tilde{\nu}\partial_a b + 6\frac{I_3}{s}\partial_s b \\
& \quad - \left(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\right)\frac{\partial_a s}{s}\Bigg]\frac{\varepsilon^2 z^2}{s}H \\
& + \Bigg[-\partial_s G + \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}}{a}b + \frac{\tilde{\nu}^2}{a^2} - \frac{\rho\tilde{\nu}}{a}\Gamma_0 - 2\rho\tilde{\nu}\partial_a b \\
& \quad + 6\frac{I_3}{s}\partial_s b - \left(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\right)\frac{\partial_a s}{s}\Bigg]\varepsilon^2 H.
\end{aligned}$$

A.5.1 Computing $\tilde{c}$

From the above equation it is clear that we should choose $\tilde{c}$ in order to cancel the $\frac{\varepsilon^2 z^2}{s}H$ term. This means we set

$$s\partial_s\tilde{c} + 2\tilde{c} - 2\rho\tilde{\nu}\partial_a b + 6\frac{I_3}{s}\partial_s b - \left(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\right)\frac{\partial_a s}{s} = 0.$$

Multiplying by s and summarizing the terms yields

$$\begin{aligned}
\partial_s\big(s^2\tilde{c}\big) + 6I_3\partial_s b &= 2s\rho\tilde{\nu}\partial_a b + \partial_a s\frac{\tilde{\nu}^2}{a} + \Big(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b\Big)\partial_a s \\
&= 2s\rho\tilde{\nu}\partial_a b + \partial_a s\frac{\tilde{\nu}^2}{a} + \Big(2\rho\tilde{\nu}\partial_s\big(sb\big) - 3\rho\tilde{\nu}b\Big)\partial_a s.
\end{aligned}$$

Taking into consideration that I_3 was chosen to satisfy

$$\partial_s I_3 = \frac{1}{2}\rho\tilde{\nu}\partial_a s,$$

we can conclude that c must solve

$$\partial_s\big(s^2\tilde{c}\big) = -6\partial_s\big(I_3 b\big) + \frac{\tilde{\nu}^2}{a}\partial_a s + 2\rho\tilde{\nu}\Big(\partial_s\big(sb\big)\,\partial_a s + s\partial_a b\Big).$$

To explicitly solve this, let us first consider the last term of this equation. We get

$$\begin{aligned}\partial_s\big(sb\big)\,\partial_a s + s\partial_a b &= \partial_s\left(\frac{1}{a}I_1\right)\partial_a s + \partial_a\left(\frac{1}{a}I_1\right)\\ &= \frac{1}{a}\partial_a S\partial_s I_1 - \frac{1}{a^2}I_1 + \frac{1}{a}\partial_a I_1. \end{aligned} \tag{A.9}$$

Recall that $I_1(s,t,a)$ is defined as

$$I_1(s,t,a) = \rho\int_{\tilde{t_0}(s,t,a)}^{t} \nu\big(y(u,t,a)\big)y(u,t,a)X(t,u,y(u,t,a))\,du.$$

In particular, this means for the a-derivative that we have

$$\begin{aligned}\partial_a I_1(s,t,a) = &-\partial_a\tilde{t_0}(s,t,a)\rho\tilde{\nu}(\tilde{t_0}(s,t,a),t,a)y^2(\tilde{t_0}(s,t,a),t,a)\\ &+\rho\int_{\tilde{t_0}(s,t,a)}^{t}\partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\,du.\end{aligned}$$

Following [16] we can see that we have

$$\partial_a S(\tilde{t_0}(s,t,a),t,a) = y^2(\tilde{t_0}(s,t,a),t,a)\partial_a\tilde{t_0}(s,t,a).$$

Thus, (A.9) reduces to

$$\begin{aligned}&\partial_s\big(sb\big)\,\partial_a s + s\partial_a b\\ &\quad= -\frac{1}{a^2}I_1(s,t,a) + \frac{1}{a}\rho\int_{\tilde{t_0}(s,t,a)}^{t}\partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\,du\\ &\quad= -\frac{1}{a}sb + \frac{1}{a}\rho\int_{\tilde{t_0}(s,t,a)}^{t}\partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\,du,\end{aligned}$$

and yields

$$\begin{aligned}&2\rho\tilde{\nu}\Big[\partial_s\big(sb\big)\,\partial_a s + s\partial_a b\Big] =\\ &\quad-\partial_s\big(s^2b^2\big) + \frac{2\rho\tilde{\nu}}{a}\rho\int_{\tilde{t_0}(s,t,a)}^{t}\partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\,du.\end{aligned}$$

Switching from $\tilde{c}$ to $c = \tilde{c} + 2b^2$ we finally get

$$\begin{aligned}\partial_s\big(s^2 c\big) = -6\partial_s\big(I_3 b\big) + \frac{\tilde{\nu}^2}{a}\partial_a s + \partial_s\big(s^2 b^2\big) \\ + \frac{2\rho\tilde{\nu}}{a}\rho\int_{\tilde{t_0}(s,t,a)}^{t} \partial_a\Big(\nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a))\Big)\,du.\end{aligned}$$

Now, we only need to integrate over s to conclude that

$$s^2 c(s,a) = s^2 b^2 + \frac{1}{a}I_2(s,t,a) - 6I_3(s,t,a)b + \frac{2}{a}I_4(s,t,a).$$

Here the integral functions are given by

$$\begin{aligned}&I_2(s,t,a) \\ &\quad= \int_0^s \partial_a S\big(\tilde{t_0}(x),t,a\big)\tilde{\nu}\big(\tilde{t_0}(x),t,a\big)^2\,dx \\ &\quad= \int_{\tilde{t_0}(s,t,a)}^{t} \partial_a S\big(u,t,a\big)\nu(y(u,t,a))^2 X(t,u,y(u,t,a))^2\,du \\ &\quad= 2\int_{\tilde{t_0}(s,t,a)}^{t} \nu(y(u,t,a))^2 X(t,u,y(u,t,a))^2 \int_u^t y(v,t,a)\partial_a y(v,t,a)\,dv\,du \\ &\quad= 2\int_{\tilde{t_0}(s,t,a)}^{t} \nu(y(u,t,a))^2 X(t,u,y(u,t,a))^2 \int_u^t y(v,t,a)X(t,v,y(v,t,a))^{-1}\,dv\,du\end{aligned}$$

and

$$\begin{aligned}&I_4(s,t,a) \\ &\quad= \int_0^s \rho\tilde{\nu}(\tilde{t_0}(x,t,a),t,a)\rho\int_{\tilde{t_0}(x,t,a)}^{t} \partial_a\Big(\nu(y(v,t,a))y(v,t,a)X(t,v,y(v,t,a))\Big)\,dv\,dx \\ &\quad= \rho^2\int_{\tilde{t_0}(s,t,a)}^{t} \psi(u,t,a)\int_u^t \partial_a\psi(v,t,a)\,dv\,du.\end{aligned}$$

Here the function ψ is defined as

$$\psi(u,t,a) = \nu(y(u,t,a))y(u,t,a)X(t,u,y(u,t,a)).$$

A.5.2 Computing G

Analogously to determining $\tilde{c}$, we set G to cancel the remaining $\varepsilon^2 H$ term. This means

$$\begin{aligned}\partial_s G &= \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}b}{a} + \frac{\tilde{\nu}^2}{a^2} - \frac{\rho\tilde{\nu}}{a}\Gamma_0 - 2\rho\tilde{\nu}\partial_a b + 6\frac{I_3}{s}\partial_s b \\ &\quad - \left(2\rho\tilde{\nu}s\partial_s b - \rho\tilde{\nu}b + \frac{\tilde{\nu}^2}{a}\right)\frac{\partial_a s}{s} \\ &= \tilde{c} + 2b^2 - 4\frac{\rho\tilde{\nu}b}{a} + \frac{\tilde{\nu}^2}{a^2} - \frac{\rho\tilde{\nu}}{a}\Gamma_0 - s\partial_s\tilde{c} - 2\tilde{c} \\ &= -\partial_s(sc) - \partial_s(sb)\Gamma_0 + \frac{\tilde{\nu}^2}{a^2}.\end{aligned}$$

Now we can integrate over s and conclude that

$$G(s,a) = -sc - sb\Gamma_0 + \frac{I_5(s,t,a)}{a^2},$$

with

$$I_5(s,t,a) = \int_0^s \tilde{\nu}(\tilde{t_0}(x,t,a),t,a)^2\,dx = \int_{\tilde{t_0}(s,t,a)}^t \nu(y(v,t,a))^2 X(t,u,y(u,t,a))^2\,du.$$

A.6 The Effective Forward Equation

With these choices we have found the following PDE for H:

$$\begin{cases}\partial_s H = \frac{1}{2}\partial_{zz}H - \frac{1}{2}\varepsilon\Gamma(\varepsilon z)\partial_z H - \varepsilon\rho\tilde{\nu}\Big(\partial_a s\,\partial_{zs}H + \partial_{za}H\Big) \\ \qquad\quad + \frac{1}{2}\varepsilon^2\tilde{\nu}^2\Big(\partial_a s^2\,\partial_{ss}H + \partial_{aa}s\,\partial_s H\Big) \\ H \to \delta(z) \text{ as } s \to 0^+,\end{cases}$$

where we have abbreviated $H(s,z,a,t,F)$ as H. Since this is the same equation as for $\tilde{Q}^{(0)}(s,z,a,t,F)$, we can identify $H(s,z,a,t,F)$ as $\tilde{Q}^{(0)}(s,z,a,t,F)$ to order $\mathcal{O}(\varepsilon^2)$. So, we have

$$H(s,z,a,t,F) = \tilde{Q}^{(0)}(s,z,a,t,F).$$

Thus, we have shown that

$$\tilde{Q}^{(2)}(s,z,a,t,F) = \tilde{Q}^{(0)}(s,z,a,t,F)e^{2\varepsilon b(s,a)z + \varepsilon^2\tilde{c}(s,a)z^2 + \varepsilon^2 G(s,a)},$$

and by further approximating the exponential function we can conclude that

$$\tilde{Q}^{(2)}(s,z,a,t,F) = \tilde{Q}^{(0)}(s,z,a,t,F)e^{\varepsilon^2 G(s,a)}(1 + 2\varepsilon b(s,a)z + \varepsilon^2 c(s,a)^2 z^2). \quad \text{(A.10)}$$

Now, to express this in terms of the original variables α and t_0, simply recall our previous changes of variable

$$a = Y(t, t_0, \alpha)$$
$$\text{and} \qquad s = \int_{t_0}^{t} y(u, t, a)^2 \, du.$$

Furthermore, recall that y was defined as the inverse of Y. Thus, we can express $y(u, t, a)$ as

$$y(u, t, a) = y(u, t, Y(t, t_0, \alpha)) = Z(t, u, t_0, \alpha).$$

With this we can express s in terms of α as

$$s = \int_{t_0}^{t} Z(t, u, t_0, \alpha)^2 \, du.$$

Making these changes in the integral functions I_1-I_5 yields those found in Assumption III.

References

[1] Andreasen, J. and Huge, B., *Expanded Forward Volatility*, RISK Magazine, 1 (2013), pp. 101–107.

[2] Antonov, A. and Misirpashaev, T., *Markovian Projection onto a Displaced Diffusion*, International Journal of Theoretical and Applied Finance, 12 (2009), pp. 507–522.

[3] Antonov, A., Arneguy, M. and Audet, N., *Markovian Projection to a Displaced Volatility Heston Model*, Available at SSRN, (2008).

[4] Antonov, A., Konikov, M., and Spector, M., *Free Boundary SABR: Natural Extension to Negative Rates*, RISK Magazine, 4 (2015).

[5] Antonov, A., Konikov, M. and Spector M., *Mixing SABR Models for Negative Rates*, RISK Magazine, 4 (2017).

[6] Douglas, J. and Rachford, H. H. , *On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables*, Transactions of the American Mathematical Society, 82 (1956), p. 421–439.

[7] Golub, G. H. and Van Loan, C. F., *Matrix Computations*, The John Hopkins University Press, 1983,1989,1996.

[8] Hagan, P. , Lesniewski, A. and Woodward, D., *Implied Volatility Formulas for Heston Models*, Wilmott Magazine, 10 (2018), pp. 44–57.

[9] Hagan, P., Kumar, D. , Lesniewski, A. S. and Woodward, D. E., *Universal Smiles*, Wilmott Magazine, (2016).

[10] Hagan, P., Kumar, D.,Lesniewski, A. and Woodward, D., *Arbitrage-Free SABR*, Wilmott Magazine, 1 (2014), pp. 60–75.

[11] Hagan, P., Lesniewski, A. and Woodward, D., *Probability Distribution in the SABR Model of Stochastic Volatility*, in Large Deviations and Asymptotic Methods in Finance, Cham, 2015, Springer International Publishing, pp. 1–35.

[12] Hagan, P., Lesniewski, A. S. and Woodward, D. E., *Effective Media Analysis for Stochastic Volatility Models*, Wilmott Magazine, 1 (2018), pp. 46–55.

[13] Hagan, P. S., *Conservative Schemes for Solving 1D PDEs*, Available at Researchgate, (2015).

[14] Hagan, P. S and Kumar, D. and Lesniewski, A. S. and Woodward, D. E., *Arbitrage free SABR*, Not published, (2013).

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

[16] Hagan, P. S., Lesniewski, A. and Woodward, D., *Impied Volatilities for Mean Reverting SABR Models*, Available at Researchgate, (2017).

[17] in 't Hout, K. and Foulon, S., *ADI Finite Difference Schemes for Option Pricing in the Heston Model with Correlation*, Int. J. Numer. Anal. Mod., 7 (2010), pp. 303–320.

[18] Jaeckel, P. and Kahl, C., *Hyp Hyp Hooray*, Wilmott Magazine, 3 (2008), pp. 70–81.

[19] Karlsmark, M., *Four Essays in Quantitative Finance*, PhD Thesis, University of Copenhagen, (2013).

[20] Kienitz, J. and McWalter, T. and Sheppard, R., *PDE Methods for SABR*, in: Novel Methods in Computational Finance, eds. Ehrhardt, Günther, ter Maten, Mathematics in Industry, Springer, (2017), pp. 265–291.

[21] Le Floc'h F. and Kennedy G., *Finite Difference Techniques for Arbitrage-Free SABR*, Journal of Computational Finance, 20(3) (2017), pp. 51–79.

[22] Lord, R., *Fifty Shades of SABR Simulation*, ICBI Global Derivatives Trading and Risk Management Conference, (2015).

[23] Madan, D. and Seneta, E., *The Variance Gamma Model for Share Market Returns*, J. Business, 63 (1990), pp. 511–524.

[24] Madan, D.B., Carr P.P., Chang E.C., *The Variance Gamma Process and Option Pricing*, European Finance Review, 2 (1998), pp. 79–105.

[25] Merton, R., *Option Pricing when Underlying Stock Returns are Discontinuous*, Journal of Financial Economics, 3 (1976), pp. 125–144.

[26] P. Hagan, D. Kumar, A. S. Lesniewski, D. E. Woodward, *Arbitrage Free SABR*, Wilmott, (2015).

[27] Peaceman, D. W. and Rachford, H. H. , *The Numerical Solution of Parabolic and Elliptic Differential Equations*, Journal of the Society for Industrial and Applied Mathematics, 3 (1955), p. 28–41.

[28] Piterbarg, V., *Markovian Projection Method for Volatility Calibration*, Available at SSRN, (2006).

[29] Press, W. H., Vetterling, W. T., Teukolsky, S. A., and Flannery, B. P., *Numerical Recipes in C++, 2nd*, Cambridge University Press, 2002.

[30] Schoebel, R. and Zhu, J., *Stochastic Volatility with an Ornstein Uhlenbeck Process: An Extension*, European Finance Review, 3 (1999), pp. 23–46.

[31] Schoutens, W., *Lévy Process in Finance - Pricing Financial Derivatives*, Wiley Series in Probability and Stochastics, 2003.

[32] Sheppard, R., *Pricing Equity Derivatives under Stochastic Volatility: A Partial Differential Equation Approach*, MSc Dissertation, University of the Witwatersrand, Johannesburg, 2007.

[33] Stein, E. M. and Stein, J. C., *Stock Price Distribution with Stochastic Volatility: An Analytic Approach*, Review of Financial Studies, 4 (1991), pp. 727–752.