From Arbitrage to Arbitrage-Free Implied Volatilities

Lech A. Grzelak[a,b,*] Cornelis W. Oosterlee[a,c]

First Version: November 24, 2014. *This version: March 3, 2016*

Abstract

We propose a method for determining an arbitrage-free density implied by Hagan's formula [1] [8]. Our method is based on the stochastic collocation method [4, 12]. The principle is to determine a few collocation points on the implied survival distribution function (SDF) and project them on a polynomial of an arbitrage-free variable for which we choose the Gaussian variable. In this way we have equality in probability at the collocation points while the generated density is arbitrage-free. Analytic European option prices are available and the implied volatilities stay very close to those initially obtained by Hagan's formula. The proposed method is very fast and straightforward to implement as it only involves 1D Lagrange interpolation and inversion of a linear system of equations. The method is generic and may be applied to other variants or other models that generate arbitrage.

Keywords: Arbitrage-free density, Collocation Method, Orthogonal Projection.

1 Introduction

When handling a large number of market volatility quotes it is natural to express them in terms of some parametric form so that a whole range of strikes can be explained by only a few parameters. Once the parametric equation is given, one can instantly obtain volatilities by evaluating the parametric function.

For several years a market standard for volatility parameterization is the well-known Hagan formula [8] which originates from a short-maturity heat kernel expansion. Although very easy to implement, the density implied by the approximation is not always free of arbitrage, especially not for very low strikes (it becomes negative or the density does not integrate to one).

Pricing of specific financial derivatives, like Constant Maturity Swaps (CMS), relies on integration of the payoff over the density which is implied from a volatility parameterization. For these CMS products industrial practice is based on marginals which should be properly defined and arbitrage-free. In other words, these marginals cannot be based on Hagan's formula. In this article we propose an alternative.

Over the last decade a number of model improvements has been introduced like in [2] where by a one time-step finite difference approximation the SABR model was solved or in [9] where the density was arbitrage-free but the method required the numerical solution of a probability density function. Other improvements on the density were introduced in [6] and [5].

Our approach differs from the ones available in the literature as we do not seek for a better analytic expression for the implied volatilities from the SABR model but we project, by means of a coordinate transformation, the survival probability onto another stochastic variable which leads to an arbitrage-free density. The concept can be expressed as follows: assuming that Y is a random variable corresponding to a model used for parameterizing volatilities and X is a

[a]Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, the Netherlands.

[b]Pricing Model Validation, Rabobank, Graadt van Roggenweg 400, 3531 AH, Utrecht, the Netherlands.

[c]CWI - National Research Institute for Mathematics and Computer Science, Science Park 123, 1098 XG, Amsterdam, the Netherlands.

*Corresponding author. E-mail address: L.A.Grzelak@tudelft.nl.

[1]We use the wording "Hagan formula" as an abbreviation to the Hagan-Kumar-Leśniewski-Woodward model.

Electronic copy available at: http://ssrn.com/abstract=2529684

known random variable, e.g. a Gaussian, we determine a coordinate transformation $y = g(x)$ for which European call prices (and their implied volatilities) agree, i.e.:

$$\int (y - K_i)^+ f_Y(y)\mathrm{d}y = \int (g(x) - K_i)^+ f_X(x)\mathrm{d}x, \quad i \in \{1, \ldots, N\}.$$

When the coordinate transformation is known, one can use it for pricing any plain vanilla product while benefiting from the fact that the density $f_X(x)$ specified is free of arbitrage. In short, the proposed method can be used to approximate a random variable Y by a polynomial based on normal variables (or another variable), i.e.,

$$Y \sim a_0 + a_1 X + a_2 X^2 + a_3 X^3 + \ldots,$$

where the coefficients $a_0, a_1, a_2, \ldots$ are inferred from a mapping (explained in Section 2.2) and by solving a small system of equations.

Our preferred method for determining the mapping relies on the stochastic collocation method [4, 12]. A mapping of a *problematic* variable Y to a variable X using only a few inversions of the CDF of Y has been presented in the financial context in [7], where a large number of Monte Carlo samples were generated based on only a few inversions of the original distribution. We start with the implied volatilities obtained by using Hagan's formula and the density function governed by that.

This article consists of a few sections. In Section 2 the application of the stochastic collocation method, projection of a "not properly defined variable" on a well-defined stochastic variable is presented. In Section 3 we discuss different projection approaches based on the collocation method and show that the stochastic collocation method enables us to determine analytic European option prices. Section 4 contains numerical experiments with well-known parameter sets for the SABR model, and Section 5 concludes.

2 Basics of stochastic collocation and implied density

2.1 Basics of stochastic collocation

Let us start with some intuition behind the collocation method. The method is developed to approximate an *expensive to compute* stochastic variable Y by means of a *cheap* variable X. An approximation is made based on the inversion of the CDF of Y at only a small set of collocation points, being the zeros of an orthogonal polynomial based on variable X.

The stochastic collocation method can be used to approximate a cumulative distribution function (CDF). As any CDF is uniformly distributed, we have $F_Y(Y) \overset{\mathrm{d}}{=} F_X(X)$. This equality in distribution does not imply that X and Y are equal in distribution, but only that the CDFs follow the same uniform distribution. From the representation above, realizations of Y, y_n, and X, x_n, are connected via the following inversion relation,

$$y_n = F_Y^{-1}(F_X(x_n)). \tag{2.1}$$

The objective is now to determine an alternative relation which does not require many of the "expensive" inversions $F_Y^{-1}(\cdot)$ for all samples of X. The task is to find a function $g(\cdot) = F_Y^{-1}(F_X(\cdot))$ such that

$$F_X(x) = F_Y(g(x)) \quad \text{and} \quad Y \overset{\mathrm{d}}{=} g(X),$$

where evaluations of function $g(\cdot)$ do not require the inversions $F_Y^{-1}(\cdot)$. With a mapping $g(\cdot)$ determined, the CDFs $F_X(x)$ and $F_Y(g(x))$ are not only equal in distributional sense but also element-wise.

Sampling from random variable Y can be decomposed into sampling from a cheap random variable X and a transformation to Y via $g(\cdot)$, i.e., $y_n = g(x_n)$. It is important to choose $g(\cdot)$ to be a simple, basic function.

To find a proper mapping function we need to extract some information from Y.

An efficient method for sampling from variable Y in terms of variable X is obtained by defining $g(\cdot)$ to be a *polynomial expansion*, i.e.

$$y_n \approx g_N(x_n) = \sum_{i=1}^{N} y_i \ell_i(x_n), \quad \ell_i(x_n) = \prod_{j=1, i\neq j}^{N} \frac{x_n - \bar{x}_j}{\bar{x}_i - \bar{x}_j}, \tag{2.2}$$

Electronic copy available at: http://ssrn.com/abstract=2529684

where x_n is a sample from X and $\bar{x}_i$ and $\bar{x}_j$ are so-called *collocation points*, y_i is the exact evaluation at collocation point $\bar{x}_i$ in (2.1), i.e., $y_i = F_Y^{-1}(F_X(\bar{x}_i))$ in (2.2). A particular choice for the collocation points $\bar{x}_i$ will be discussed in Section 3.

2.2 Implied density

The Stochastic Alpha Beta Rho (SABR) model from [8] is described by the following system of stochastic differential equations:

$$\begin{aligned} \mathrm{d}S(t) &= \sigma(t)S^{\beta}(t)\mathrm{d}W_1(t), \quad S(t_0) = S_0, \\ \mathrm{d}\sigma(t) &= \gamma\sigma(t)[\rho\mathrm{d}W_1(t) + \sqrt{1-\rho^2}\mathrm{d}W_2(t)], \quad \sigma(t_0) = \alpha \end{aligned}$$

where $S(t)$ is the forward rate, S_0 the initial forward rate, σ represents the stochastic volatility, and with the parameters ρ, β, γ, α denoting, respectively, the correlation, the skew, the volatility of volatility (vol-vol) and the overall level of the volatility parameters. The model is popular in the financial industry because of the availability of an analytic asymptotic implied volatility formula (derived with the help of perturbation theory), which is, for completeness, presented in Appendix A.

This implied volatility parameterization formula is often used in the financial industry for expressing the market quotes, even for options with expiry times of twenty years or more. It is however a well-known fact that the accuracy of this so-called Hagan formula deteriorates with time and so the occurrence of implied densities giving rise to arbitrage opportunities increases as the option expiry times increase.

In this section, we discuss how to determine an arbitrage-free density based on $f_Y(\cdot)$ of Y. Typically, the implied density has problems around 0 where the absorption property is not properly handled in the formula[2]. The density deteriorates in a region near zero (see Fig. 1, left-side picture). We will map Y onto a random variable X, such that the mapping procedure takes place in those regions where the density of Y is properly defined.

The representation in (2.2) with $y_i = F_Y^{-1}(F_X(\bar{x}_i))$ is not yet well-suited. The main problem comes from the fact that the implied CDF does not have the natural $[0, 1]$ bounds, as shown in Figure 1 (right-side picture). Since the density can become negative, CDF $F_Y(y)$ exhibits an upper bound which is less than one.

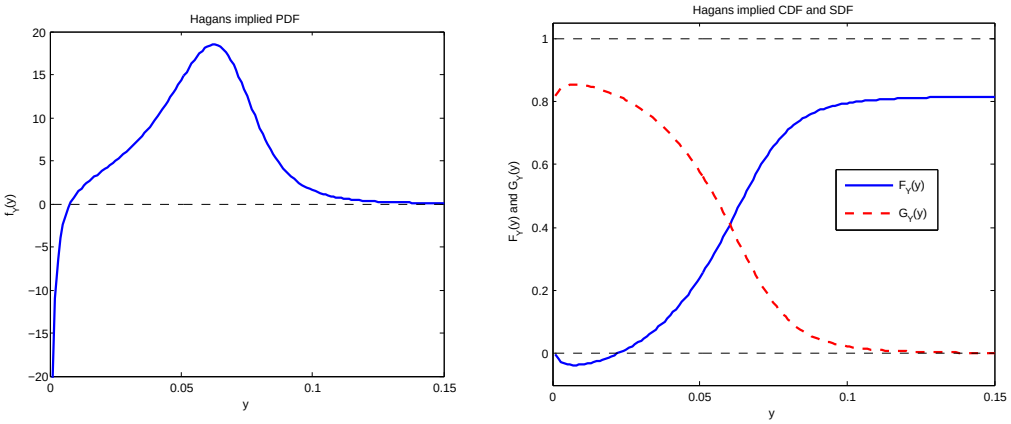

Figure 1: parameter values $\beta = 0.5$, $\alpha = 0.05$, $\rho = -0.7$, $\gamma = 0.4$, $F(t_0) = 0.05$ and $T = 7$. Left: probability density, with deterioration near zero; right: corresponding CDF (blue) and SDF (dotted red).

Since $F_Y(\cdot)$ is not well-defined the inversion $F_Y^{-1}(F_X(\bar{x}_i))$ will give us incorrect mapping points. Figure 1 shows however that, although $F_Y(y)$ does not have proper upper and lower bounds, the *survival distribution function* (SDF), defined by means of the European call options, $V_{\text{call}}(t_0, K)$, with strike K, as

$$G_Y(y) = 1 - \int_{-\infty}^{y} f_Y(x)\mathrm{d}x = \int_{y}^{+\infty} f_Y(x)\mathrm{d}x = \boxed{-\frac{\partial V_{\text{call}}(t_0, K)}{\partial K}\Big|_{K=y}}, \tag{2.3}$$

[2]By using the so-called "shifted variant of the model", i.e. based on $S + \theta$, see [9], the instability at zero shifts towards $S = \theta$.

has as its natural limit value[3] of 0 for $y \to \infty$. By focussing on the survival distribution $G_Y(\cdot)$, we can make use of the collocation mapping, which is given by:

$$y \approx g_N(x) = \sum_{i=1}^{N} G_Y^{-1}(G_X(\bar{x}_i))\ell_i(x) =: \sum_{i=1}^{N} y_i \ell_i(x), \quad \ell_i(x) \text{ in (2.2).} \tag{2.4}$$

Remark (Proper calculation of the survival distribution)**.** *To determine the survival probability it is crucial not to use the representation $G_Y(y) = 1 - F_Y(y)$ but to integrate from y to infinity, as presented at the right-hand side of (2.3). By integration we essentially find the zero region and ensure that the SDF has a proper limit value, under the assumption that the right-side tail converges to zero.*

Because the inversion $G_Y^{-1}(G_X(\bar{x}_i))$ is only well-defined in the part of $G_Y(\cdot)$ which is monotone, we set specific values $g_{\min}$ and $g_{\max}$ and choose the collocation points so that $G_X(\bar{x}_1) < g_{\max}$ and $G_X(\bar{x}_N) > g_{\min}$. In other words, the values of $g_{\min}$ and $g_{\max}$ determine the range at which we can be confident about the quality of the mapping between the two variables.

When the limits $g_{\min}$ and $g_{\max}$ are prescribed, the collocation method maps the survival probability of Y onto a survival probability based on a polynomial of X. In Figure 2 an intuitive schematic picture of the projection routine is presented. Note that the Lagrange interpolation (right upper figure) takes place at the variable level (X, Y), which is typically rather smooth and almost linear (see [7] for a more detailed discussion). In order to apply Lagrange interpolation between the nodes $(\bar{x}_i, y_i)$ it is important to make use of optimal collocation points $\bar{x}_i$ ensuring that the polynomial has certain optimality properties, and avoiding any oscillations to occur. In Section 3 the details regarding the choice of these optimal points $\bar{x}_i$ will be given.

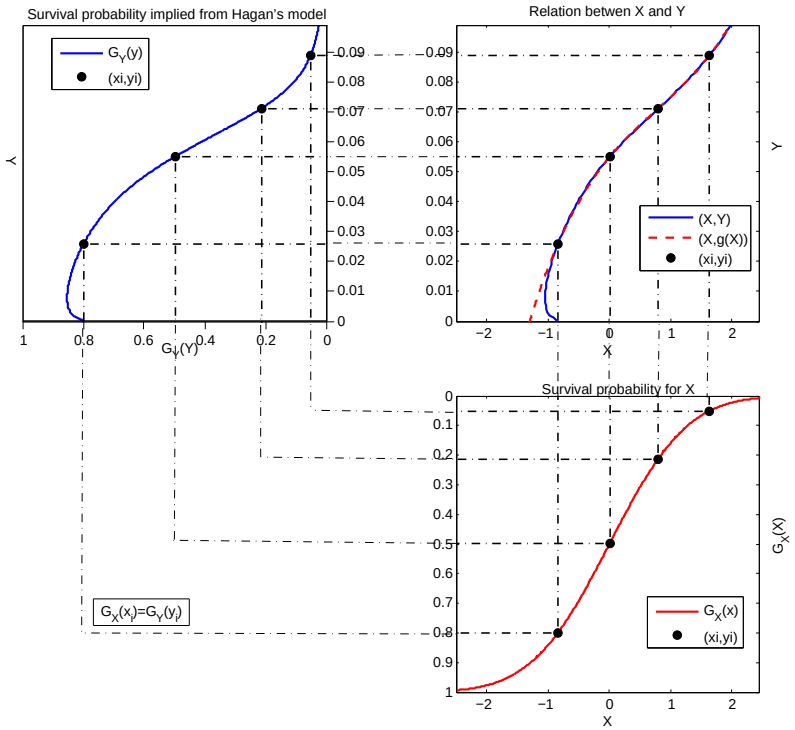

Figure 2: Illustration of the mappings of Y on $X \sim \mathcal{N}(0,1)$ with a polynomial $g_N(X)$. The parameters are taken as in Section 4.1.

In the next subsection we will discuss the relation between the densities of Y and X.

[3]We could use that $\exists \hat{y}\ G_Y(\hat{y}) = 0.5$ so that $F_Y(\hat{y}) = 0.5$ to determine an adjustment factor for the CDF.

2.3 Recovery of the PDF and pricing options

By the definition of function $g(x)$, we have:

$$y = g(x) \overset{\text{def}}{=} G_Y^{-1}(G_X(x)), \quad \text{so} \quad G_X(x) = G_Y(g(x)) =: G_Y(y). \tag{2.5}$$

Differentiating (2.5) w.r.t x results in:

$$\frac{\mathrm{d}G_X(x)}{\mathrm{d}x} = -f_X(x), \qquad \frac{\mathrm{d}G_Y(g(x))}{\mathrm{d}x} = -f_Y(g(x))\frac{\mathrm{d}g(x)}{\mathrm{d}x},$$

and the relation between the densities is therefore given by:

$$f_Y(g(x)) = f_X(x)\left(\frac{\mathrm{d}g(x)}{\mathrm{d}x}\right)^{-1} \approx f_X(x)\left(\frac{\mathrm{d}g_N(x)}{\mathrm{d}x}\right)^{-1}, \tag{2.6}$$

with $g_N(x)$ as in (2.4) and derivative $\mathrm{d}g_N(x)/\mathrm{d}x$ is known analytically, see below in (2.7).

By means of the mapping $y \to x$, we can make use of the *cheap* PDF of X, $f_X(\cdot)$. In order to evaluate $f_X(x)$, however, the mapping $x \approx g_N^{-1}(y)$ needs to be calculated.

Remark (Efficient evaluation of $x = g^{-1}(y)$)**.** *Since mapping $y = g(x)$ is bijective and $g(x)$ is strictly increasing, so is $g^{-1}(y)$. This implies that the arguments x can be obtained by the inverse interpolation [10] of $g(x)$ against y, which can be done at essentially no cost.*

With $g_N(x)$ the Lagrange polynomial, its derivative reads:

$$\frac{\mathrm{d}g_N(x)}{\mathrm{d}x} = \sum_{i=1}^{N} y_i \frac{\mathrm{d}\ell_i(x)}{\mathrm{d}x} = \sum_{i=1}^{N} y_i \ell_i(x) \sum_{j=1, j\neq i}^{N} \frac{1}{x - \bar{x}_j}, \tag{2.7}$$

and the density can be further simplified to:

$$f_Y(g(x)) \approx f_X(x)\left(\sum_{i=1}^{N} y_i \ell_i(x) \sum_{j=1, j\neq i}^{N} \frac{1}{x - \bar{x}_j}\right)^{-1}. \tag{2.8}$$

Lemma 2.1 (Arbitrage-free collocation variable)**.** *The probability density function as given in (2.6) is free of arbitrage, and integrates to 1.*

Proof. The proof is straightforward: by integrating the density $f_Y(y)$, which is not well-defined, and changing variables $y = g(x)$, we find:

$$\int_{-\infty}^{+\infty} f_X(x)\left(\frac{\mathrm{d}g_N(x)}{\mathrm{d}x}\right)^{-1} \mathrm{d}g_N(x) = 1. \tag{2.9}$$

The proper limits of the corresponding functions $G_X(x)$ and $F_X(x)$ are guaranteed by the choice of X. ■

Using the results above, we can price European-style payoffs highly efficiently, as:

$$\begin{aligned} V(t_0, y_0) &= \mathbb{E}\left[V(T, Y(T))|\mathcal{F}(t_0)\right] = \int_0^{\infty} V(T, y) f_Y(y)\mathrm{d}y \\ &= \int_{G_X^{-1}(1)}^{G_X^{-1}(0)} V(T, g(x)) f_Y(g(x)) \frac{\mathrm{d}g(x)}{\mathrm{d}x}\mathrm{d}x, \end{aligned} \tag{2.10}$$

which, by Equation (2.6) and the approximation $g(x) \approx g_N(x)$, gives:

$$V(t_0, y_0) = \int_{G_X^{-1}(1)}^{G_X^{-1}(0)} V(T, g(x)) f_X(x)\mathrm{d}x \approx \int_{G_X^{-1}(1)}^{G_X^{-1}(0)} V(T, g_N(x)) f_X(x)\mathrm{d}x. \tag{2.11}$$

Although the pricing of options is generally done numerically, by integrating expression in (2.11), European put and call option prices are known analytically when X is a Gaussian variable (see Section 3.4).

3 Details of the stochastic collocation

We start with a discussion how to choose a proper set of collocation points. With these collocation points determined, we can derive analytic European option prices. As a last item of this section, we discuss a re-calibration procedure for the exact fitting of implied volatilities.

3.1 Collocation points determined by the strikes

We first consider the case in which, for a given set of strikes $K = (y_1, y_2, \ldots, y_N)^{\mathrm{T}}$, a calibration with the Hagan formula needs to be performed. At this point we assume that the corresponding SDF is monotone in the interval $[y_1, y_N]$, or, in other words, that the SDF is *well-defined* in that region. Typically, the strikes at which the calibration then takes place are *not too close* to the zero region where the implied density may not be arbitrage-free. At the set of collocation points $(\bar{x}_1, \ldots, \bar{x}_N)^T$ the survival probabilities implied by the model and by the collocation variable X will be identical.

We take the collocation variable to be a standard normally distributed variable, $X \sim \mathcal{N}(0,1)$, the coordinate transformation is then given by Equation (2.4) with $y_i = G_Y^{-1}(G_{\mathcal{N}(0,1)}(\bar{x}_i))$, with G_Y as in (2.3). We determine the collocation points $\bar{x}_i$ at which we have:

$$G_Y^{-1}(G_{\mathcal{N}(0,1)}(\bar{x}_i)) = y_i, \quad \Rightarrow \quad \bar{x}_i = G^{-1}_{\mathcal{N}(0,1)}(G_Y(y_i)). \tag{3.1}$$

With a set of strike values (i.e. points y_i), we only need to determine the corresponding survival probabilities $G_Y(y_i)$ and use them in an inversion procedure based on a standard normal variable, which is trivial [4].

When the collocation points $\bar{x}_i$ are given by means of the specified strikes $K_i = y_i$, the constructed Lagrange polynomial may *not be stable* for *large values of* N ($N \geq 5$). For large values of N, it is recommended to use the quadrature points as presented in Section 3.2 where the use of collocation points preserves interpolation stability.

3.2 Collocation points obtained by quadrature

An alternative to choosing collocation points according to the strikes available in the market is to determine the collocation points $\bar{x}_i$ in an *optimal manner* based on the zeros of an orthogonal polynomial. As explained in [7], since we choose X to be normally distributed the $\bar{x}_i$ points will be the zeros of the Hermite polynomials [1].

However, we need to make sure that the collocation points do not require inversions $G_Y^{-1}(G_X(\bar{x}_i))$ at points where $G_Y(\cdot)$ is not well-defined. Therefore we will apply a so-called *grid stretching technique*, as introduced in [7], and define a stretched variable $Z := a + bX$, with $X \sim \mathcal{N}(0,1)$. We specify two parameters $g_{\min}$ and $g_{\max}$ to define the range in which the inversion can be safely performed. We subsequently determine values a and b so that $G_Z(\bar{x}_1) = g_{\max}$, $G_Z(\bar{x}_N) = g_{\min}$, which implies that $1 - F_{\mathcal{N}(a,b)}(\bar{x}_1) = g_{\max}$ and $1 - F_{\mathcal{N}(a,b)}(\bar{x}_N) = g_{\min}$. After some basic algebraic operations, we arrive at:

$$b = \frac{\bar{x}_1 - \bar{x}_N}{F^{-1}_{\mathcal{N}(0,1)}(1 - g_{\max}) - F^{-1}_{\mathcal{N}(0,1)}(1 - g_{\min})}, \quad a = \bar{x}_1 - bF^{-1}_{\mathcal{N}(0,1)}(1 - g_{\max}). \tag{3.2}$$

So, $G_Z(\bar{x}_i) \in [g_{\min}, g_{\max}]$ and the following y_i values will be obtained:

$$y_i = G_Y^{-1}\left(G_{\mathcal{N}(0,1)}\left(\frac{\bar{x}_i - a}{b}\right)\right), \tag{3.3}$$

where $\bar{x}_i$ are the *optimal* collocation points of $X \sim \mathcal{N}(0,1)$. These points are available based on the moments of a standard normal variable (see [7] for the details).

Note that Gauss quadrature has the optimal polynomial degree for integration. By the grid stretching approach described above, one may loose the theoretical optimality properties. However, grid stretching maintains the stability of interpolation, as it only leads to a linear shift of quadrature nodes.

[4] Note that $G^{-1}_{\mathcal{N}(0,1)}(\cdot) = -F^{-1}_{\mathcal{N}(0,1)}(\cdot)$, with $F_{\mathcal{N}(0,1)}(\cdot)$ is the CDF of a standard normal variable.

3.3 Absorption at zero

A consequence of using a polynomial based on a normally distributed random variable X for the approximation of variable Y is that an arbitrage-free density may give rise to a positive probability of *negative values*, i.e., $\mathbb{P}[g(X) < 0] > 0$.

An *absorption* at zero boundary condition can be easily incorporated into the methodology by the following constraint on the function $g_N(X)$:

$$\widehat{g}_N(X) = \begin{cases} g_N(X), & \text{for} \quad g_N(X) > 0 \Leftrightarrow X > g_N^{-1}(0), \\ 0, & \text{for} \quad g_N(X) \leq 0 \Leftrightarrow X \leq g_N^{-1}(0), \end{cases} \tag{3.4}$$

with $X > g_N^{-1}(0)$ corresponding to the condition $Y > 0$, with $Y \approx g_N(X)$.

Polynomial $\widehat{g}_N(X)$ has an atom at $X = g_N^{-1}(0)$ which corresponds to an atom at $Y = 0$. The corresponding probability mass is given by:

$$\mathbb{P}[Y = 0] \approx \mathbb{P}[\widehat{g}_N(X) = 0] = \mathbb{P}[g_N(X) \leq 0].$$

The absorption boundary condition in (3.4) assigns the probability of $g_N(X)$ becoming negative to a one point at 0.

For $K > 0$ both polynomials, $g_N(X)$ and $\widehat{g}_N(X)$, will yield the same European option prices. This can be seen by prescribing $V(T, Y(T)) = \max(Y(T) - K, 0)$ in (2.11) which then yields:

$$\max(Y(T) - K, 0) \quad \approx \quad \max(\widehat{g}_N(X) - K, 0) = \max(g_N(X) - K, 0),$$

as $\widehat{g}_N(X) = g_N(X)$ for $g_N(X) > 0$. The difference between $g_N(X)$ and $\widehat{g}_N(X)$ will only be present when dealing with negative values.

In Figure 3 in Section 4.1 an illustrative example will be given.

3.4 Analytic European option prices for normal collocation variable

Before we give the analytic expression for European option prices, we recall the formulas for the moments of a truncated normal distribution.

Result 3.1 (The moments for a truncated univariate normal distribution)**.** *Let $X \sim \mathcal{N}(0,1)$ and $a \in (-\infty, +\infty)$, then the expression for the moments $m_i := \mathbb{E}[X^i | X > a]$, reads*

$$m_i = (i-1)m_{i-2} + \frac{a^{i-1} f_{\mathcal{N}(0,1)}(a)}{1 - F_{\mathcal{N}(0,1)}(a)}, \quad i = 1, \ldots, \tag{3.5}$$

with $m_{-1} = 0$, $m_0 = 1$ and $f_{\mathcal{N}(0,1)}(x)$ and $F_{\mathcal{N}(0,1)}(x)$ the standard normal probability and cumulative distribution functions, respectively.

In the following lemma we show that European option prices under $g_N(X)$ with $X \sim \mathcal{N}(0,1)$ are known analytically.

Lemma 3.1 (European call option prices)**.** *With the collocation random variable $X \sim \mathcal{N}(0,1)$ for $g_N(X)$, European call prices are analytically available, and given by:*

$$V_{call}(t_0, K) \approx \int_{G_X^{-1}(1)}^{G_X^{-1}(0)} (g_N(x) - K)^+ f_{\mathcal{N}(0,1)}(x) \mathrm{d}x = G_{\mathcal{N}(0,1)}(c_K) \left[\sum_{i=0}^{N-1} a_i \mathbb{E}[X^i | X > c_K] - K \right],$$

with $c_K = g_N^{-1}(K)$, $G_{\mathcal{N}(0,1)}(c_K) = 1 - F_{\mathcal{N}(0,1)}(c_K)$, $\mathbb{E}[X^i | X > c_k]$ the moments of the truncated normal variable, given in Result 3.1 and where a_i, $i \in \{0, \ldots, N-1\}$, are (constant) coefficients obtained by inverting Vandermonde matrix, V, in the matrix equation $V\boldsymbol{a} = \boldsymbol{y}$. The k'th row of matrix V is given by $(1, \bar{x}_k^1, \bar{x}_k^2, \ldots, \bar{x}_k^{N-1})$, with $\bar{x}_i$ the predetermined collocation points.

Proof. Integration domain $(G_X^{-1}(1), G_X^{-1}(0))$ for a normally distributed random variable X is given by $(-\infty, +\infty)$. A Lagrange function $g_N(x)$ can be expressed in terms of monomials, $\mathbf{m}(x) = (1, x, x^2, \ldots, x^{N-1})^{\mathrm{T}}$, as:

$$g_N(x) = a_0 + a_1 x + \cdots + a_{N-1} x^{N-1}, \quad \text{with} \quad g_N(\bar{x}_i) = y_i, \tag{3.6}$$

with coefficients a_i, $i \in \{0, \ldots, N-1\}$. The coefficients $\mathbf{a} = (a_0, \ldots, a_{N-1})^{\mathrm{T}}$ are determined by the interpolation conditions, $g_N(\bar{x}_i) = y_i$, for $i = 1, \ldots, N$. These coefficients can be found as solutions of linear system $V\boldsymbol{a} = \boldsymbol{y}$, i.e.,

$$\begin{pmatrix} 1 & \bar{x}_1^1 & \bar{x}_1^2 & \ldots & \bar{x}_1^{N-1} \\ 1 & \bar{x}_2^1 & \bar{x}_2^2 & \ldots & \bar{x}_2^{N-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \bar{x}_{N-1}^1 & \bar{x}_{N-1}^2 & \ldots & \bar{x}_{N-1}^{N-1} \\ 1 & \bar{x}_N^1 & \bar{x}_N^2 & \ldots & \bar{x}_N^{N-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-2} \\ a_{N-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \\ y_N \end{pmatrix}, \tag{3.7}$$

with matrix V the Vandermonde matrix.

Given the monomial representation of function $g_N(x)$, we set $c_K = g_N^{-1}(K)$ for which the option pricing equation becomes:

$$\int_{-\infty}^{+\infty} \max(g_N(x) - K, 0) f_{\mathcal{N}(0,1)}(x)\mathrm{d}x = \int_{c_K}^{+\infty} g_N(x) f_{\mathcal{N}(0,1)}(x)\mathrm{d}x - K\mathbb{P}[X > c_K].$$

By definition of $g_N(x)$ in (3.6) we have:

$$\begin{aligned} \int_{c_K}^{+\infty} g_N(x) f_{\mathcal{N}(0,1)}(x)\mathrm{d}x - K\mathbb{P}[X > c_K] &= \sum_{i=0}^{N-1} a_i \int_{c_K}^{+\infty} x^i f_{\mathcal{N}(0,1)}(x)\mathrm{d}x - K\mathbb{P}[X > c_K] \\ &= \sum_{i=0}^{N-1} a_i \mathbb{E}[X^i 1_{X > c_K}] - K\mathbb{P}[X > c_K], \end{aligned} \tag{3.8}$$

where $X \sim \mathcal{N}(0,1)$.

The last integral defines $\mathbb{E}[X^i 1_{X>c_K}]$, which is equivalent to:

$$\begin{aligned} \mathbb{E}[X^i 1_{X>c_K}] &= \mathbb{E}[X^i | X > c_K]\mathbb{P}[X > c_K] \\ &= \mathbb{E}[X^i | X > c_K](1 - F_{\mathcal{N}(0,1)}(c_K)). \end{aligned} \tag{3.9}$$

By combining (3.9) and (3.8) the proof is complete. ■

By the put-call parity and the lemma above, put option prices are also available in closed form. Moreover, with analytic European option prices the calculation of the corresponding implied volatility is a trivial exercise.

Remark (Inversion of the Vandermonde matrix)**.** *In order to calculate these European option prices, one needs to invert the Vandermonde matrix in (3.7). For a large number of collocation points ($N > 8$), this can be problematic and therefore it is important to check the condition number of this matrix.*

3.5 Method enhancement and re-calibration

We cannot guarantee that the arbitrage-free density obtained by stochastic collocation yields the same implied volatilities as those given by Hagan's formula. Here, we address this issue.

Let $\Omega_Y = [\beta, \alpha, \rho, \gamma]$ be the set of parameters which is initially obtained by calibrating with Hagan's formula to find the implied volatilities, $\sigma_Y^{\Omega}(\overline{K})$, with $\overline{K}$ the set of market strike values. This set of parameters will thus give us density $f_Y^{\Omega}(y)$, which may not be free of arbitrage. By the collocation method we subsequently obtain a density $\widehat{f}_X^{\Omega}(x)$ which is arbitrage-free but does not necessarily return the implied volatilities for which $\sigma_X^{\Omega}(\overline{K}) = \sigma_Y^{\Omega}(\overline{K})$.

In essence, function $g(\cdot)$ does not only depend on X, but is also a function of the parameters, i.e. $g(X; \Omega)$, so, by changing Ω we may obtain a different set of option prices and corresponding volatilities. Therefore, we define here an *optimization problem* in which we search for $\widehat{\Omega}$ so that $g(X; \widehat{\Omega})$ gives *the same implied volatilities* as those from the market.

In other words, we describe an optimization procedure in which we determine the set of parameters, $\widehat{\Omega} = [\widehat{\beta}, \widehat{\alpha}, \widehat{\rho}, \widehat{\gamma}]$, so that the volatilities from the market and the collocation method agree, i.e.

$$\min_{\widehat{\Omega}} \sum_{\overline{K}} \left(\sigma_{Mrkt}(\overline{K}) - \sigma_{g(X,\widehat{\Omega})}(\overline{K}) \right)^2. \tag{3.10}$$

Typically $\widehat{\Omega}$ will be *close* to the initial set Ω and a local-search algorithm, like the Nelder-Mead [11] method, will therefore be sufficient to find the optimal solution.

The optimization problem in (3.10) can also be expressed in terms of European option prices. The gradients required during optimization in that case may be derived analytically.

Additionally, one may include constraints on the coefficients of polynomial $g_N(X)$ ensuring for example the martingale property of the model, i.e. $\mathbb{E}[g_N(X)] = S_0$, with S_0 the forward price. Since $g_N(X)$ is given as a polynomial, we have:

$$\mathbb{E}[g_N(X)] = \sum_{i=0}^{N-1} a_i \mathbb{E}[X^i] = S_0, \tag{3.11}$$

which e.g. with $N = 4$ and $X \sim \mathcal{N}(0,1)$ gives us $S_0 = a_0 + a_2$.

4 Numerical experiments

We discuss several numerical experiments, ranging from a simple illustrative example to testing the introduced method with model parameters that are often used in the literature and considered to be *extreme and realistic*.

4.1 Illustrative numerical experiment

We continue with the experiment in Section 2.2, with the following set of parameters: $\beta = 0.5$, $\alpha = 0.05$, $\rho = -0.7$, $\gamma = 0.4$, $F(t_0) = 0.05$ and $T = 7$.

For the implied survival probabilities we take $g_{\max} = 0.8$ and $g_{\min} = 0.05$, implying that the density from the parameterization is well-defined in the range $[G_Y^{-1}(g_{\max}), G_Y^{-1}(g_{\min})]$. We choose $N = 4$ collocation points. Since we project the model on a polynomial of normal variables, we first determine the optimal collocation points, $\bar{x}_i$, for $X \sim \mathcal{N}(0,1)$. These points are well-known as they are the zeros [5] of the Hermite polynomials:

$$\left(\bar{x}_1^0, \bar{x}_2^0, \bar{x}_3^0, \bar{x}_4^0\right) = (-2.3344, -0.7420, 0.7420, 2.3344)\,.$$

For details on how to obtain these points, we refer to [7]. Using Equation (3.2), $a = -0.7541$ and $b = 1.8777$ and by $\bar{x}_i = (\bar{x}_0 - a)/b$, the collocation points are given by:

$$(\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4) = (-0.8416, 0.0065, 0.7968, 1.6448)\,.$$

For each collocation point we calculate, see Equation (3.3), $y_i = G_Y^{-1}(G_X(\bar{x}_i))$:

$$(y_1, y_2, y_3, y_4) = (0.0258, 0.0551, 0.0713, 0.0894)\,.$$

Points y_i should be considered as *optimal* points at which the survival probabilities are mapped. It is easy to check that $G_X(\bar{x}_1) = G_Y(y_1) = g_{\max} = 0.8$ and $G_X(\bar{x}_4) = G_Y(y_4) = g_{\min} = 0.05$. In Figure 2 the mapping procedure is illustrated and in Figure 3 the absorption technique is shown.

With collocation points $\bar{x}_i$ and nodes y_i, the mapping is $y \approx g_4(x) = \sum_{i=1}^4 y_i \ell_i(x)$, the price of a European call option with strike K is given by Lemma 3.1.

In Figure 3 the recovered SDFs and the relation between the variables X and Y are presented. The stochastic collocation method guarantees an exact match at the collocation points which are indicated by the black dots. Between the collocation points the variables are interpolated by Lagrange interpolation and outside the range of collocation points an extrapolation, determined by X, takes place.

From a computational perspective the collocation method is highly efficient. The complete simulation takes about [6] $0.01s$, which is highly satisfactory.

In the considered example the application of the collocation method gave highly satisfactory results already without the re-calibration step. It will be shown in the next section that the results can be further improved.

4.2 Market examples

Here, we test our method with parameters that are well-known from the literature. Different parameter combinations are presented in Table 1, where the option expiry varies from $1y$ to $15y$. In the experiments we show the generated densities and the corresponding implied

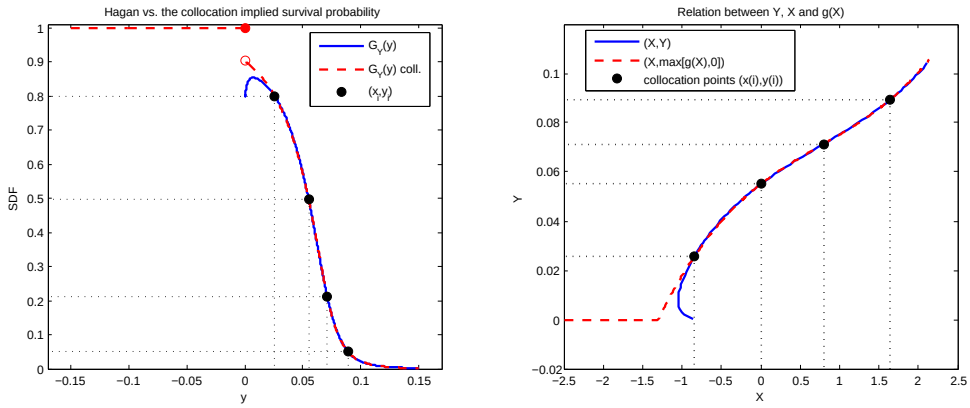

Figure 3: Left: SDFs obtained from the implied density and the collocation method. Right: relation between (X, Y) and $(X, \widehat{g}_N(X))$, as given in (3.4). The parameters are set as $\beta = 0.5$, $\alpha = 0.05$, $\rho = -0.7$, $\gamma = 0.4$, $F(t_0) = 0.05$ and $T = 7$.

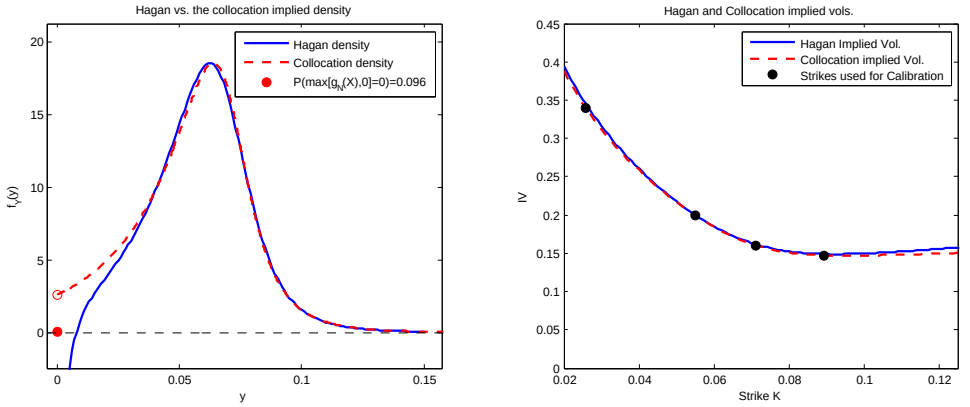

Figure 4: Left: The implied density and the density by the collocation method. Right: The corresponding implied volatilities. The parameters are as in Figure 3.

Parameters:	β	α (ATM)	ρ (Corr)	γ (vol-vol)	$F(t_0)$	T
Set I as in [3]	0.6	0.25	-0.8	0.3	1	10
Set II as in [9]	0.25	0.35	-0.1	1	1	1
Set III as in [5]	0.2	0.26	-0.5	0.35	1	15

Table 1: Model parameters chosen in the experiments.

volatilities. In all experiments we consider four collocation points that are determined based on the strikes $y_i = K_i$, as presented in Section 4.1. The method thus employs only four collocation points to reproduce the implied volatilities from the market.

In Figures 5, 6 and 7 the obtained survival probabilities (SDF) and the implied volatilities are presented. In all three cases the SDFs from the collocation method are as desired, i.e., they are monotone and their limits are 0 and 1. The resulting implied volatilities are not all very close to the market values. This can be improved by performing the re-calibration step. In all three cases the re-calibration results in an almost perfect implied volatility match at the collocation points. We also note that the tail asymptotics and the level of curvature and skewness were preserved by the stochastic collocation method. With as few as four collocation points, in all examples, a wide range of implied volatility shapes were generated.

Because the re-calibration step requires only local optimization iterations, it is very fast. The full projection and calibration procedure takes less than 0.1 seconds for all three cases.

[5]Scaled by $\sqrt{2}$.

[6]i5-4670 CPU @ 3.40GHz with 8.00GB ram simulated with Matlab

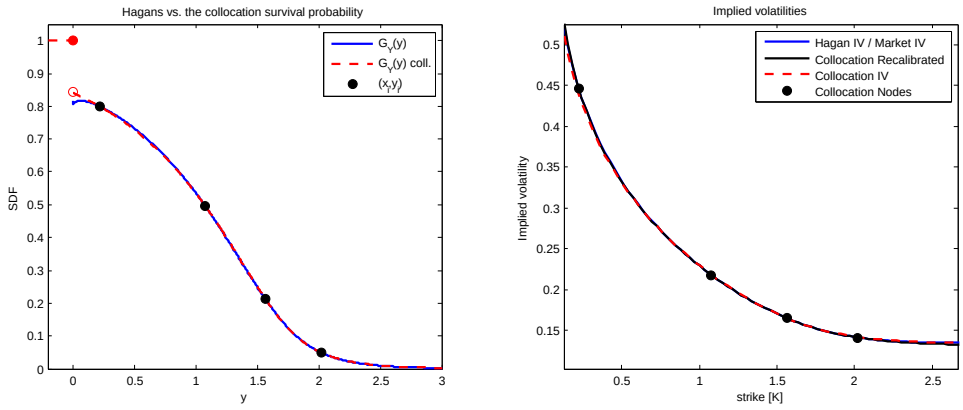

Figure 5: Survival probabilities and implied volatilities for Set I (as given in [3]) with and without re-calibration. The experiment was performed with $g_{\min} = 0.05$ and $g_{\max} = 0.8$. $G_Y(y)$ stands for the implied survival probability while "$G_Y(y)coll$" indicates the survival probability obtained from the collocation method.

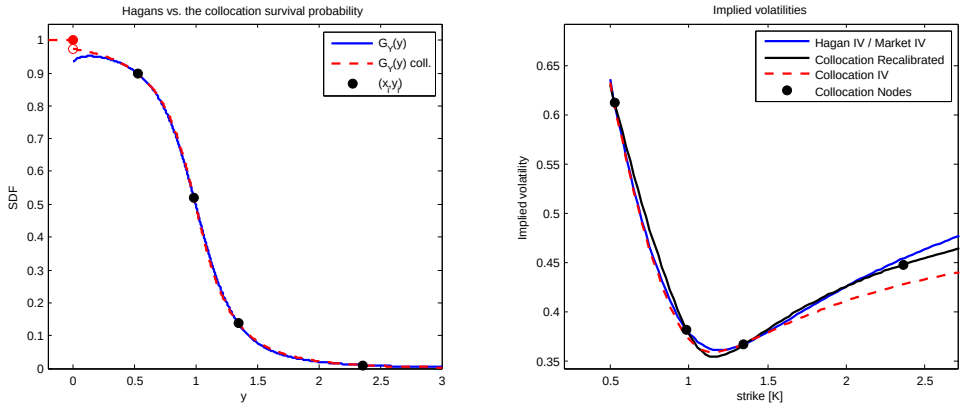

Figure 6: Survival probabilities and implied volatilities for Set II (as given in [9]) with and without re-calibration. The experiment was performed with $g_{\min} = 0.01$ and $g_{\max} = 0.9$.

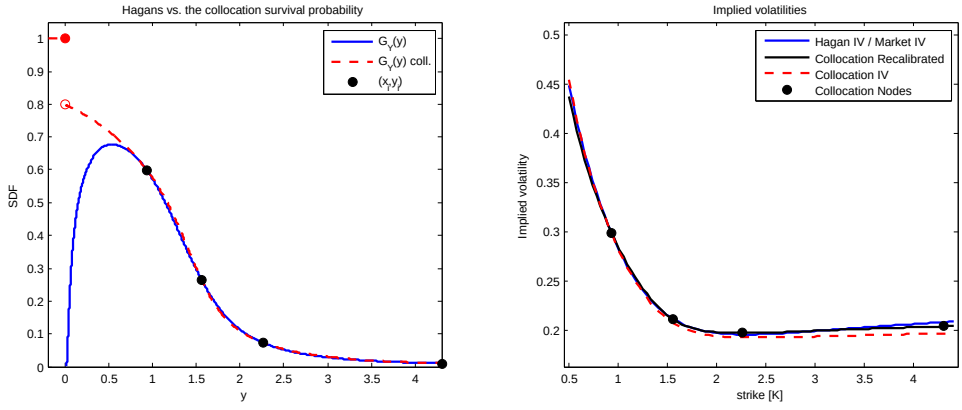

Figure 7: Survival probabilities and implied volatilities for Set III (as given in [5]) with and without re-calibration. The experiment was performed with $g_{\min} = 0.01$ and $g_{\max} = 0.6$.

5 Conclusions

In this article we have presented an application of the stochastic collocation method for obtaining an arbitrage-free density based on Hagan's formula. Our method relies on the availability of a survival distribution function, which is not necessarily well-defined on the whole domain, and it is projected onto a Gaussian variable. The method presented gives implied volatilities in accordance with those obtained by the model, however, in some cases a re-calibration step is required to guarantee a *perfect* fit. The method is easy to implement as it only relies on Lagrange interpolation and the solution of a linear system of equations.

References

[1] M. Abramowitz and I.A. Stegun. *Modified Bessel Functions* **I** *and* **K**. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th edition, 1972.

[2] J. Andreasen and B. Huge. Expanded forward volatility. *Risk*, pages 101–107, 2013.

[3] A. Antonov and M. Spector. Advanced analytics for the SABR model. *SSRN*, 2012.

[4] I. Babuška, F. Nobile, and R. Tempone. A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, 45:1005–1034, 2007.

[5] P. Balland and Q. Tran. SABR goes normal. *Risk*, pages 76–81, 2013.

[6] P. Doust. No-arbitrage SABR. *The Journal of Computational Finance*, 15(3):3–31, 2012.

[7] L.A. Grzelak, J.A.S. Witteveen, M. Suarez, and C.W. Oosterlee. The stochastic collocation Monte Carlo sampler: Highly efficient sampling from "expensive" distributions. *Available at SSRN: http://ssrn.com/abstract=2529691*, 2014.

[8] P.S. Hagan, D. Kumar, A.S. Leśniewski, and D.E Woodward. Managing smile risk. *Wilmott Magazine*, pages 84–108, 2002.

[9] P.S. Hagan, D. Kumar, A.S. Leśniewski, and D.E. Woodward. Arbitrage-free SABR. *Wilmott Magazine*, pages 60–75, 2014.

[10] E. Michalup. On inverse linear interpolation. *Scandinavian Actuarial Journal*, pages 98–100, 1950.

[11] J. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[12] D. Xiu and J.S. Hesthaven. High-order collocation methods for differential equations with random inputs. *SIAM Journal on Scientific Computing*, 27:1118–1139, 2005.

A Implied volatility for the SABR model

The approximating implied volatility derived in [8] reads:

$$\hat{\sigma}^H(T,K) \quad = \quad A(K)\frac{z(K)}{\chi(z(K))} + B(T,K), \quad \text{and} \quad \hat{\sigma}^H(T,S_0) = \frac{\alpha}{S_0^{1-\beta}}B(T,S_0),$$

where

$$\begin{aligned}
z(K) &= \frac{\gamma}{\alpha}(S_0K)^{(1-\beta)/2}\log(S_0/K), \quad \chi(z(K)) = \log\left(\frac{\sqrt{1-2\rho z(K)+z^2(K)}+z(K)-\rho}{1-\rho}\right),\\
A(K) &= \alpha\left(S_0K^{(1-\beta)/2}\left(1+\frac{(1-\beta)^2}{24}\log^2(S_0/K)+\frac{(1-\beta)^4}{1920}\log^4(S_0/K)\right)\right)^{-1},\\
B(T,K) &= \left\{1+\left(\frac{(1-\beta)^2}{24}\frac{\alpha^2}{(S_0K)^{1-\beta}}+\frac{1}{4}\frac{\rho\beta\gamma\alpha}{(S_0K)^{(1-\beta)/2}}+\frac{2-3\rho^2}{24}\gamma^2\right)\right\}T.
\end{aligned}$$