Approximate Arbitrage-Free Option Pricing under the SABR Model

Nian Yang[a,*], Nan Chen[b], Yanchu Liu[c], Xiangwei Wan[d]

[a]*Department of Finance and Insurance, Nanjing University, 22 Hankou Road, Nanjing, 210093, China.*
[b]*Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.*
[c]*Lingnan (University) College, Sun Yat-sen University, Guangzhou, 510275, China*
[d]*Antai College of Economics and Management, Shanghai Jiao Tong University, 1954 Huashan Road, Shanghai, 200030, China.*

Abstract

The stochastic-alpha-beta-rho (SABR) model introduced by Hagan et al. (2002) provides a popular vehicle to model the implied volatilities in the interest rate and foreign exchange markets. To exclude arbitrage opportunities, we need to specify an absorbing boundary at zero for this model, which the existing analytical approaches to pricing derivatives under the SABR model typically ignore. This paper develops closed-form approximations to the prices of vanilla options to incorporate the effect of such a boundary condition. Different from the traditional normal distribution-based approximations, our method stems from an expansion around a one-dimensional Bessel process. Extensive numerical experiments demonstrate its accuracy and efficiency. Furthermore, the explicit expression yielded from our method is appealing from the practical perspective because it can lead to fast calibration, pricing, and hedging.

Keywords: SABR model; Approximate solution; Arbitrage-free option pricing; Perturbation method

JEL classification: C63 G13

*Corresponding author.
Email addresses: yangnian@nju.edu.cn (Nian Yang), nchen@se.cuhk.edu.hk (Nan Chen), liuych26@mail.sysu.edu.cn (Yanchu Liu), xwwan@sjtu.edu.cn (Xiangwei Wan)

1. Introduction

The SABR model introduced by Hagan et al. (2002) is widely used in the interest rate and foreign exchange markets. It is a local stochastic volatility model in which the underlying (forward price or rate) follows a constant elasticity of variance (CEV)-type diffusion process and the dynamics of volatility is governed by a geometric Brownian motion. The popularity of the SABR model comes from its closed-form asymptotic implied volatility formula and capturing the correct comovement between the smile dynamics and the forward price (see, e.g., Hagan et al., 2002).

However, the forward price under the SABR model may hit zero with positive probability, resulting in arbitrage opportunities. Therefore an absorbing boundary condition at zero must be specified to avoid arbitrage opportunities (see, e.g., Delbaen and Shirakawa, 2002; Rebonato et al., 2009). A number of analytical approaches presented in the existing literature ignore this requirement when they attempt to develop approximate solutions from the pricing partial differential equations (PDEs). This leads to that some extensively used pricing formulas, such as those given by Hagan et al. (2002), Obłój (2008) and Paulot (2015), could not rule out arbitrage opportunities. Our research aims to address the issue how to derive new approximations for the vanilla option price under the SABR model to incorporate the effect of the absorbing boundary condition.

To this end, this paper introduces a novel combination of three transformations to exploit the structural properties of the SABR model. They are rescaling, the Lamperti transformation, and homogenization. These transformations result in two expansion parameters based on the total volatility of volatility (vol-of-vol) and the correlation between the underlying price and the volatility. The leading order operator in the transformed pricing PDE is the infinitesimal generator of a one-dimensional Bessel process, from which we can derive an explicit zero-order approximation. Furthermore, the transformations we use reveal clearly that the above leading order solution will achieve a second-order accuracy because no first-order differential operators for either the total vol-of-vol or the correlation appear in the SABR model after the transformations. To the best of our knowledge, our approximation takes a new approach relative to the exisiting literature. We also carry out an exhaustive set of numerical experiments to demonstrate the accuracy and efficiency of our method.

Our work relates to three strands of literature. First, a variety of ad-hoc numerical methods have been proposed to address the absorbing boundary at zero for the SABR model. Nevertheless, none of these progress presents (approximate) analytical formulas for the option price. For instance, Doust (2012) numerically calculates the probability that the forward price hits zero. In the presence of an absorbing/reflecting boundary at zero, Hagan et al. (2015) analyze the model in a special case that the correlation is zero. They also introduce the normal SABR model, which is further analyzed by Balland and Tran (2013). Hagan et al. (2014) propose a numerical scheme to solve a PDE with an absorbing boundary simplified from the SABR model. Gulisashvili et al. (2015) study the probability of hitting zero without correlation and further apply it to small strike implied volatilities. Compared with this literature, our paper is the first one to derive closed-form approximation for the SABR model with an absorbing boundary and to analyze its accuracy order.

The second strand of the literature is on pricing continuously monitored barrier options. The traditional analytical methods are often based on the symmetric structure of the underlying models. Shreve (2004), Jiang (2005) apply the "reflection principle" or the "method of image" to obtain analytical formulas for barrier options under the Black-Scholes model. Davydov and Linetsky (2001), Linetsky (2007) develop a spectral method to tackle general one-dimensional models. All one-dimensional diffusion models in these literature possess some symmetry properties since their infinitesimal generators are at least formally self-adjoint (see, e.g., formula (3.2) of Linetsky (2007)). Kwok et al. (1998) provide analytical formulas for barrier options under the multivariate Black-Scholes model, using the symmetric structures of this model.

The third strand of literature is on asymptotic expansions of the option prices, which are popular in the option valuation because of their efficiency and flexibility. One well-know method is based on the theory for analyzing generalized Wiener functionals initiated by Watanabe (1987). For more details and references, one can refer to Li (2013) and Li (2014) for its applications to statistical inference and vanilla option valuation, respectively. Another attractive method based on perturbations of PDEs can be used to find asymptotic formulas for both vanilla and exotic options. See, e.g., Fouque et al. (2000), Widdicks et al. (2005) for an application to vanilla options. As for barrier options, Howison and Steinberg (2007) and Ilhan et al. (2004) derive asymptotic formulas for barrier options under the Black-Scholes model and a fast mean-reverting stochastic volatility

model, respectively, mainly leveraging on the symmetric structures of these two models.

Under the SABR model, it turns out that pricing a vanilla call without arbitrage is equivalent to pricing a down-and-out call with a knock-out boundary at zero. However, the SABR model is not symmetric, which makes the aforementioned approaches invalid. The paper contributes to the literature by proposing a new method via a transformation combination to tame the intractability.

The remainder of the paper proceeds as follows: In Section 2, we present the SABR model and the formulation of the arbitrage-free option pricing problem. In Section 3, we derive the analytical formulas to approximate arbitrage-free vanilla option prices. In Section 4, we numerically justify the validity of our formulas through option prices and implied volatilities. This paper is concluded in Section 5. The related proofs and derivations are collected in the Appendixes.

2. The SABR Model and Problem Formulation

2.1. The SABR Model

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathbb{P}$ is a T-forward martingale measure (cf. Section 9.6.2 of Musiela and Rutkowski (2004) or Section 9.4 of Shreve (2004)). There are two independent Brownian motions $\{B_t; 0 \le t \le T\}$ and $\{W_t; 0 \le t \le T\}$ defined on it. Let $\{\mathcal{F}_t^B; 0 \le t \le T\}$ and $\{\mathcal{F}_t^W; 0 \le t \le T\}$ be the information filtrations generated by the two Brownian motions respectively. Define $\mathcal{F}_t = \mathcal{F}_t^B \otimes \mathcal{F}_t^W$. Denote F_t and A_t to be the underlying asset's forward price and volatility at time t, for any arbitrary $t \in [0, T]$, respectively. The SABR model is specified by the following system of stochastic differential equations (SDEs):

$$\begin{cases} dF_t = A_t F_t^{\beta}[\sqrt{1-\rho^2} dB_t + \rho dW_t], \\ dA_t = \nu A_t dW_t, \end{cases} \tag{1}$$

where $\beta \in [0, 1)$,[1] $\nu > 0$, $\rho \in (-1, 1)$. The initial points F_0 and A_0 are positive. The parameter ν is known as the volatility of volatility, which plays an important role in the following expansion for the SABR model. It is clear that this is a local stochastic volatility model in which the forward

[1] If $\beta = 1$, the forward price is a lognormal process, which is always positive. This trivial case is not considered throughout this paper.

price $F = \{F_t; 0 \le t \le T\}$ follows a CEV-type diffusion process and the dynamics of the volatility $A = \{A_t; 0 \le t \le T\}$ is given by a geometric Brownian motion.

The process F can hit zero with positive probability. A reflecting boundary is not appropriate because it will lead to an arbitrage opportunity: we can buy the forward at zero cost when F hits 0 and sell it for profit when it is reflected back to the positive region. To avoid the arbitrage opportunity, we must impose an absorbing boundary condition for F at the origin; that is, if F hits 0, it will remain there from then on. One can refer to Rebonato et al. (2009) for a detailed discussion. Henceforth, we impose the following assumption on the model:

Assumption 1. 0 *is an absorbing boundary of the process* F.

This assumption not only ensures that there is no arbitrage opportunity but also ensures the uniqueness of the solution to the SDE (1).

2.2. *Formulation of the Arbitrage Free Option Pricing Problem*

Under the T-forward measure, pricing the derivative can be achieved by calculating an expectation of its payoff (see, e.g., formula (2.20) of Brigo and Mercurio (2006)). Therefore, the arbitrage-free European option price $V_h(t, f, a)$ at time t is given by the following conditional expectation:

$$V_h(t, f, a) = \mathbb{E}[h(F_T)|F_t = f, A_t = a], \tag{2}$$

where $h(\cdot)$ is the payoff function, which can be a call with payoff function $h(f) = (f - K)^+$ or a put with $h(f) = (K - f)^+$, and K is the strike price. The corresponding option prices are $V_c(t, f, a)$ and $V_p(t, f, a)$, respectively. Let

$$\tau_t := \min\{s \ge t : F_s = 0\} \tag{3}$$

be the first time the process F hits the lower boundary 0, to which we refer as the barrier hereafter. Under the assumption that the forward price F does not hit zero before time t, (2) can be rewritten as follows:

$$V_h(t, f, a) = \mathbb{E}[h(F_T)\mathbf{1}_{\{\tau_t > T\}} + h(0)\mathbf{1}_{\{\tau_t \le T\}}|F_t = f, A_t = a].$$

Specifically, if it is a call option, i.e., the payoff function is $h(f) = (f - K)^+$, then

$$V_c(t, f, a) = \mathbb{E}[(F_T - K)^+ \mathbf{1}_{\{\tau_t > T\}} | F_t = f, A_t = a]. \tag{4}$$

Hence, pricing a European call under the SABR model without arbitrage is equivalent to pricing a down-and-out call option with a knock-out boundary at zero. If it is a put option, then

$$V_p(t, f, a) = \mathbb{E}[(K - F_T)^+ \mathbf{1}_{\{\tau_t > T\}} | F_t = f, A_t = a] + K \cdot \mathbb{E}[\mathbf{1}_{\{\tau_t \le T\}} | F_t = f, A_t = a]. \tag{5}$$

Thus, pricing a put is essentially equivalent to pricing a rebate option. The rebate option expires when the forward price reaches zero, at which time and the option holder is refunded with a certain premium (the amount is the probability of hitting zero). The same formulation can be found in Hagan et al. (2014).

We first focus on deriving an approximate formula for the price of the call option without arbitrage, then the put is addressed similarly. The arbitrage-free European call option price $V_c(t, f, a)$, providing sufficient smoothness, is the solution to a backward Kolmogorov PDE with boundary and terminal conditions. The PDE for $V_c(t, f, a)$ is specified in the theorem below.

Theorem 1. *Assume that $\varphi(t, f, a)$ is differentiable in t, twice differentiable in f and a in the interior of $[0, T] \times \mathbb{R}^+ \times \mathbb{R}^+$. $\varphi(t, f, a)$ is continuous to the boundary, bounded in a and at most linear growth in f. Moreover, for $t \in [0, T)$ and $f, a \in (0, +\infty)$, $\varphi(t, f, a)$ satisfies the following backward Kolmogorov PDE*

$$\frac{\partial \varphi}{\partial t} + \frac{1}{2}\left(a^2 f^{2\beta} \frac{\partial^2 \varphi}{\partial f^2} + 2\rho\nu a^2 f^{\beta} \frac{\partial^2 \varphi}{\partial f \partial a} + \nu^2 a^2 \frac{\partial^2 \varphi}{\partial a^2}\right) = 0, \tag{6}$$

with boundary and terminal conditions

$$\varphi(t, 0, a) = 0, \quad \varphi(T, f, a) = h(f). \tag{7}$$

Then, $\varphi(t, f, a)$ admits a stochastic representation as follows:

$$\varphi(t, f, a) = \mathbb{E}[h(F_T) \mathbf{1}_{\{\tau_t > T\}} | F_t = f, A_t = a].$$

In particular, such a solution is unique.

Proof. See Appendix A. □

3. Approximate Solutions to the Arbitrage-Free Option Pricing Problem

In this section, we shall develop approximate formulas for the arbitrage-free price of vanilla options. We first consider the problem of pricing the call, i.e., solving the PDE (6) with the payoff function being fixed at $h(f) = (f - K)^+$ in (7). As shown in Section 3.1, three transformations on (6) change the problem into a new one (19). Note that the leading order differential operator in (19) is the corresponding infinitesimal generator of a one-dimensional Bessel process. This feature results in a lot of tractability when we solve for the zero-order approximation. More importantly, by exploiting the structure of the SBAR model, the three transformations we used also ensure that all the first-order terms disappear when we expand the equation in terms of the total vol-of-vol and the correlation. In other words, the zero-order approximation we obtain in this section can actually achieve a high-order accuracy, providing a theoretical foundation to explain why our approximation works very well in the numerical experiments, as illustrated in the subsequent section. We derive the approximate formula for the put option price in Section 3.3.

3.1. Transformations

As aforementioned, the key leading to our approximation is three transformation steps that we shall discuss in details in this subsection. We first *rescale* the time and state variables (cf. (8)) to introduce a new perturbation parameter, the total vol-of-vol. Then, we carry out the *Lamperti transformation* (11) to unitize the volatility coefficient into 1. This operation yields a new differential equation (13) whose leading order differential operator $\mathcal{L}_0$ is the infinitesimal generator of a one-dimensional Bessel process. Finally, we use a *homogenization* procedure to remove the total vol-of-vol in the initial (terminal) condition, accomplishing the last step to an analytical zero-order approximation.

3.1.1. Rescaling

We rescale the time and volatility variables while keeping the forward price unchanged as follows:

$$\tau = \frac{T - t}{T}, \quad f = f, \quad g = \frac{a}{\nu}. \tag{8}$$

Let $C_1(\tau, f, g)$ be the call option price after rescaling; that is,

$$C_1(\tau, f, g) := V_c(t, f, a) \equiv V_c(T(1-\tau), f, \nu g). \tag{9}$$

Obviously, the following relationships should hold between the derivatives of functions $V_c(t, f, a)$ and $C_1(\tau, f, g)$:

$$\frac{\partial V_c}{\partial t} = (-\frac{1}{T})\frac{\partial C_1}{\partial \tau}, \quad \frac{\partial V_c}{\partial a} = \frac{1}{\nu}\frac{\partial C_1}{\partial g}, \quad \frac{\partial^2 V_c}{\partial a^2} = \frac{1}{\nu^2}\frac{\partial^2 C_1}{\partial g^2}, \quad \frac{\partial^2 V_c}{\partial f \partial a} = \frac{1}{\nu}\frac{\partial^2 C_1}{\partial f \partial g}.$$

Substituting the above derivatives into the PDE (6), the new function $C_1(\tau, f, g)$ in (9) under the new coordinates (τ, f, g) now satisfies the following equation:

$$\frac{\partial C_1}{\partial \tau} = \frac{\epsilon^2}{2}\left(g^2 f^{2\beta}\frac{\partial^2 C_1}{\partial f^2} + 2\rho g^2 f^{\beta}\frac{\partial^2 C_1}{\partial f \partial g} + g^2\frac{\partial^2 C_1}{\partial g^2}\right), \; C_1(\tau, 0, g) = 0, \; C_1(0, f, g) = (f-K)^+. \tag{10}$$

Note that a new parameter $\epsilon = \nu\sqrt{T}$ arises in (10). We refer to it as the total vol-of-vol from now on.

3.1.2. Lamperti Transformation

The Lamperti transformation is a standard technique in the literature to transform a one-dimensional diffusion process with a general volatility coefficient into a new diffusion with volatility of unity (that is, the coefficient of the diffusion term is 1). As one of the major innovative techniques of this paper, we apply it to the SABR model, an example of two-dimensional diffusions. Define new coordinates (x, y) as follows:

$$x = \int_0^f \frac{du}{\epsilon u^{\beta} g} = \frac{f^{1-\beta}}{\epsilon(1-\beta)g}, \qquad y = g. \tag{11}$$

By it, we should have

$$\begin{cases} \frac{\partial x}{\partial f} = \frac{1}{\epsilon f^{\beta} g}, \\ \frac{\partial x}{\partial g} = \frac{-x}{g}, \\ \frac{\partial^2 x}{\partial f^2} = \frac{-\beta}{\epsilon f^{1+\beta} g}, \\ \frac{\partial^2 x}{\partial f \partial g} = \frac{-1}{\epsilon f^{\beta} g^2}, \\ \frac{\partial^2 x}{\partial g^2} = \frac{2x}{g^2}; \end{cases} \qquad \begin{cases} \frac{\partial}{\partial f} = \frac{\partial x}{\partial f}\frac{\partial}{\partial x}, \\ \frac{\partial^2}{\partial f^2} = (\frac{\partial x}{\partial f})^2\frac{\partial^2}{\partial x^2} + \frac{\partial^2 x}{\partial f^2}\frac{\partial}{\partial x}, \\ \frac{\partial}{\partial g} = \frac{\partial x}{\partial g}\frac{\partial}{\partial x} + \frac{\partial}{\partial y}, \\ \frac{\partial^2}{\partial g^2} = (\frac{\partial x}{\partial g})^2\frac{\partial^2}{\partial x^2} + 2(\frac{\partial x}{\partial g})\frac{\partial^2}{\partial x \partial y} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2 x}{\partial g^2}\frac{\partial}{\partial x}, \\ \frac{\partial^2}{\partial f \partial g} = (\frac{\partial x}{\partial f}\frac{\partial x}{\partial g})\frac{\partial^2}{\partial x^2} + (\frac{\partial x}{\partial f})\frac{\partial^2}{\partial x \partial y} + \frac{\partial^2 x}{\partial f \partial g}\frac{\partial}{\partial x}. \end{cases}$$

Let $C_2(\tau, x, y)$ denote the function $C_1(\tau, f, g)$ transformed under the new coordinates (x, y); that is,

$$C_2(\tau, x, y) := C_1(\tau, f, g) \equiv C_1(\tau, (\epsilon(1-\beta)xy)^{1/(1-\beta)}, y). \tag{12}$$

For $\tau \in (0, 1]$, $x > 0$, simple calculations show that the transformation changes the PDE (10) regarding C_1 into the following PDE regarding C_2:

$$\mathcal{L}_0 C_2(\tau, x, y) = (\epsilon\rho\mathcal{L}_1 + \epsilon^2\mathcal{L}_2)C_2(\tau, x, y),\ C_2(\tau, 0, y) = 0,\ C_2(0, x, y) = ((\epsilon(1-\beta)xy)^{2\theta} - K)^+, \tag{13}$$

where $\theta = 1/(2(1-\beta))$ and

$$\mathcal{L}_0 = \frac{\partial}{\partial\tau} - \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1-2\theta}{2x}\frac{\partial}{\partial x},\ \mathcal{L}_1 = -x\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} + y\frac{\partial^2}{\partial x\partial y},\ \mathcal{L}_2 = \frac{1}{2}x^2\frac{\partial^2}{\partial x^2} + x\frac{\partial}{\partial x} + \frac{1}{2}y^2\frac{\partial^2}{\partial y^2} - xy\frac{\partial^2}{\partial x\partial y}. \tag{14}$$

Now we can see clearly the advantages of such transformations under the SABR model from the initial-boundary value problem (13). First, note that $\theta > 0$ since $\beta \in [0, 1)$. Via this transformed equation, we manage to relate the SABR model to the celebrated Bessel process because the leading order operator $\mathcal{L}_0$ is exactly the same as the infinitesimal generator of the latter with an absorbing boundary at 0.[2] A body of literature, upon which our zero-order approximation is based, has been developed around how to explicitly solve the initial-boundary value problem associated with $\mathcal{L}_0$; see, e.g., Davydov and Linetsky (2001); Polyanin (2001). In contrast with this existent literature on one-dimensional diffusion models, a technical barrier still remains for the SABR model due to its two-dimensional nature. In particular, the initial value $C_2(0, x, y)$ is a non-smooth function and contains a fraction order of ϵ. That entails the last step we will discuss in the next subsection. Second, the right hand side of the PDE (13) does not contain first-order terms in either ϵ or ρ. As illustrated in Section 3.2, this observation implies that the approximation error of our approach will be in the order of $O(\max(\epsilon^2, |\rho|\epsilon))$.

As far as we know, the introduction of this special Lamperti transformation is also new to the study of the SABR model. The previous research, such as Hagan et al. (2002) and Doust (2012),

[2] A special case is $\beta = 0$, in which $\mathcal{L}_0$ turns out to degenerate the corresponding infinitesimal generator of a one-dimensional Brownian motion.

often use the infinitesimal generator of a Brownian motion as the leading term of their expansion. Compared with these methods, the numerical experiments show that our new expansion can achieve higher accuracy, especially in the presence of the absorbing boundary at zero.

3.1.3. Homogenization

To overcome the difficulty caused by the fractional order of ϵ in the initial value condition, we use a further homogenization step in this subsection to remove ϵ from $C_2(0, x, y)$. Denote

$$k = \frac{K^{1-\beta}}{\epsilon(1-\beta)}. \tag{15}$$

Then we have

$$C_2(0, x, y) = ((\epsilon(1-\beta)y \cdot x)^{2\theta} - (\epsilon(1-\beta) \cdot k)^{2\theta})^+ = \gamma(\epsilon)((xy)^{2\theta} - k^{2\theta})^+, \tag{16}$$

where

$$\gamma(\epsilon) := (\epsilon(1-\beta))^{2\theta}. \tag{17}$$

Define a new function $C(\tau, x, y)$ by dividing the coefficient $\gamma(\epsilon)$ from the function $C_2(\tau, x, y)$:

$$C(\tau, x, y) := \gamma^{-1}(\epsilon)C_2(\tau, x, y) \equiv (\epsilon(1-\beta))^{-2\theta}C_1(\tau, (\epsilon(1-\beta)xy)^{2\theta}, y). \tag{18}$$

From (13), we conclude that $C(\tau, u, v)$ satisfies the following equation:

$$\mathcal{L}_0 C(\tau, x, y) = (\epsilon\rho\mathcal{L}_1 + \epsilon^2\mathcal{L}_2)C(\tau, x, y),\ C(\tau, 0, y) = 0,\ C(0, x, y) = ((xy)^{2\theta} - k^{2\theta})^+, \tag{19}$$

where $\mathcal{L}_0$, $\mathcal{L}_1$ and $\mathcal{L}_2$ are defined in (14). Note that the initial condition (cf. the last equality of (19)) does not contain the fraction order of total vol-of-vol any longer.

Our approach differs from the traditional singular perturbation method. The singular perturbation approach solves the initial-boundary value problem (see, e.g., Kevorkian and Cole, 1996; Widdicks et al., 2005) by matched asymptotic expansion. Our analysis takes a brand-new route: a combination of the Lamperti transformation (11) and the homogenization (16) transforms the original PDE to (19), in which the perturbation parameters just appear in the equation itself, neither in the boundary nor in the initial conditions.

3.2. Approximate Solution to the Arbitrage-Free Call Option Price

So far we have introduced two perturbation parameters, ϵ and ρ, in the pricing PDE (19) through three transformations. In order to derive an approximate solution to it and inspired by Fouque et al. (2003), we carry out a double asymptotic analysis in the regime where both of them are small independent parameters. In the following derivation we choose to expand first with respect to ϵ and subsequently with respect to ρ. This choice is more convenient than the reverse ordering which in fact gives the same result.

Consider the following expansion for the solution to Eq. (19), in powers of ϵ first:

$$C(\tau, x, y) = C_\rho^{(0)}(\tau, x, y) + \epsilon \cdot C_\rho^{(1)}(\tau, x, y) + O(\epsilon^2). \tag{20}$$

Substituting the RHS of (20) into Eq. (19) and comparing the coefficients of the constant term and the ϵ-term, we find that the leading term $C_\rho^{(0)}(\tau, x, y)$ and the first-order term $C_\rho^{(1)}(\tau, x, y)$ should satisfy respectively

$$\mathcal{L}_0 C_\rho^{(0)}(\tau, x, y) = 0,\ C_\rho^{(0)}(\tau, 0, y) = 0,\ C_\rho^{(0)}(0, x, y) = ((xy)^{2\theta} - k^{2\theta})^+, \tag{21}$$

and

$$\mathcal{L}_0 C_\rho^{(1)}(\tau, x, y) = \rho \mathcal{L}_1 C_\rho^{(0)}(\tau, x, y),\ C_\rho^{(1)}(\tau, 0, y) = 0,\ C_\rho^{(1)}(0, x, y) = 0. \tag{22}$$

From (21), we can see that $C_\rho^{(0)}(\tau, x, y)$ is indeed independent of ρ. Furthermore, noting that $\mathcal{L}_0$ is the corresponding infinitesimal generator of a one-dimensional Bessel process, $C_\rho^{(0)}(\tau, x, y)$ is explicitly solvable according to the following lemma.

Lemma 1. *For all continuous functions $f(\tau, x)$ and $g(x)$, which exhibit at most linear growth in x, the solution to the equation*

$$\mathcal{L}_0 P(\tau, x) = \left(\frac{\partial}{\partial \tau} - \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1 - 2\theta}{2x}\frac{\partial}{\partial x} \right) P(\tau, x) = f(\tau, x),$$

with $P(\tau, 0) = 0$ and $P(0, x) = g(x)$ can be computed from

$$P(\tau, x) = \int_0^\tau \int_0^{+\infty} \Lambda(\tau - s, x, \xi) f(s, \xi) d\xi ds + \int_0^{+\infty} \Lambda(\tau, x, \xi) g(\xi) d\xi,$$

where $\Lambda(\tau, x, \xi)$ *is given by*

$$\Lambda(\tau, x, \xi) = \frac{x^{\theta}\xi^{1-\theta}}{\tau} \exp\left(-\frac{x^2+\xi^2}{2\tau}\right) I_{\theta}\left(\frac{x\xi}{\tau}\right), \tag{23}$$

and $I_{\theta}(z)$ *is the modified Bessel function of the first kind given below:*

$$I_{\theta}(z) = \sum_{m=0}^{+\infty} \frac{(z/2)^{2m+\theta}}{m!\Gamma(1+m+\theta)}.$$

Proof. One can refer to the content in Section 1.2 of Polyanin (2001) or Section 21 in Appendix I of Borodin and Salminen (2002). □

Invoking the previous lemma, detailed calculation in Appendix B shows that

$$\begin{aligned} C_{\rho}^{(0)}(\tau, x, y) =& \int_0^{+\infty} \Lambda(\tau, x, \xi)((\xi y)^{2\theta} - k^{2\theta})^{+} d\xi \\ =& (xy)^{2\theta}\left(1 - Q\left(\frac{k^2}{\tau y^2}; 2\theta+2, \frac{x^2}{\tau}\right)\right) - k^{2\theta} Q\left(\frac{x^2}{\tau}; 2\theta, \frac{k^2}{\tau y^2}\right). \end{aligned} \tag{24}$$

where

$$Q(x; \kappa, \lambda) = \int_0^x q(\xi; \kappa, y) d\xi \text{ and } q(\xi; \kappa, y) = \frac{e^{-(\xi+y)/2}}{2}\left(\frac{\xi}{y}\right)^{(\kappa-2)/4} I_{\kappa/2-1}(\sqrt{\xi y}). \tag{25}$$

Here, $q(x; \kappa, y)$ and $Q(\xi; \kappa, y)$ are the density and cumulative distribution functions of a noncentral chi-square distribution with non-centrality y and degree κ, respectively.

Then, we further expand $C_{\rho}^{(1)}(\tau, x, y)$ with respect to ρ, i.e., $C_{\rho}^{(1)}(\tau, x, y) = C^{(1,0)}(\tau, x, y) + O(|\rho|)$, and substitute its RHS into Eq. (22). By matching coefficients we know $C^{(1,0)}(\tau, x, y)$ satisfies the following PDE, which further implies $C^{(1,0)}(\tau, x, y) = 0$ according to Lemma 1:

$$\mathcal{L}_0 C^{(1,0)}(\tau, x, y) = 0,\ C^{(1,0)}(\tau, 0, y) = 0,\ C^{(1,0)}(0, x, y) = 0.$$

That means, we actually have $C_{\rho}^{(1)}(\tau, x, y) = O(|\rho|)$ and hence the solution to Eq. (19) can be written as

$$C(\tau, x, y) = C_{\rho}^{(0)}(\tau, x, y) + O(\epsilon \cdot \max(\epsilon, |\rho|)). \tag{26}$$

Given the formula for $C_\rho^{(0)}(\tau, x, y)$ above, we can obtain the approximate formula for the call option price under the original coordinates. Recall that $\theta = 1/(2(1-\beta))$. Combining the transformations (8), (11), and (15) as well as the functions (9), (12), and (18) defined after each transformation, then we have the following approximate solution $\bar{V}_c(t, f, a)$ to the arbitrage-free call option price under the original coordinates:

$$\begin{aligned}\bar{V}_c(t,f,a) =& f \cdot \left(1 - Q\left(\frac{K^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}; \frac{3-2\beta}{1-\beta}, \frac{f^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}\right)\right) \\ & - K \cdot Q\left(\frac{f^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}; \frac{1}{1-\beta}, \frac{K^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}\right).\end{aligned} \tag{27}$$

Moreover, from (18) and (26), the approximation of the call option price now reads as follows:

$$V_c(t,f,a) = \bar{V}_c(t,f,a) + O(\epsilon^{\frac{2-\beta}{1-\beta}} \cdot \max(\epsilon, |\rho|)). \tag{28}$$

3.3. Approximate Solution to the Arbitrage-Free Put Option Price

Let $\varphi_p(t, f, a)$ and $H(t, f, a)$ be the down-and-out put option price and the probability that the forward hits zero, which are given by

$$\varphi_p(t,f,a) = \mathbb{E}[(K - F_T)^+ \mathbf{1}_{\{\tau_t > T\}} | F_t = f, A_t = a], \quad H(t,f,a) = \mathbb{E}[\mathbf{1}_{\{\tau_t \le T\}} | F_t = f, A_t = a]. \tag{29}$$

Thus by (5), the arbitrage-free put option price is given by

$$V_p(t,f,a) = \varphi_p(t,f,a) + K \cdot H(t,f,a). \tag{30}$$

Mimicking the procedure in Sections 3.1 and 3.2, we can derive an approximate formula for the price of the down-and-out put option. The only difference is that the payoff function is $h(f) = (K-f)^+$ instead of $(f-K)^+$ in PDE (10). Similarly, $\varphi_p(t, f, a)$ can be approximated by $\bar{\varphi}_p(t, f, a)$ below (A sketch of the derivation is provided in Appendix C.)

$$\begin{aligned}\varphi_p(t,f,a) =& \bar{\varphi}_p(t,f,a) + O(\epsilon^{\frac{2-\beta}{1-\beta}} \cdot \max(\epsilon, |\rho|)), \\ \bar{\varphi}_p(t,f,a) =& K \cdot \left(1 - Q\left(\frac{f^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}; \frac{1}{1-\beta}, \frac{K^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}\right)\right) \\ & - f \cdot Q\left(\frac{K^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}; \frac{3-2\beta}{1-\beta}, \frac{f^{2(1-\beta)}/(1-\beta)^2}{a^2(T-t)}\right).\end{aligned} \tag{31}$$

Moreover, according to Theorem 4.5 in Yang and Wan (2016) and its proof, the probability of hitting zero $H(t, f, a)$ in (29) has the following representation:

$$H(t, f, a) = \frac{1}{\Gamma(\theta)}\Gamma\left(\theta, \frac{f^{2(1-\beta)}/(1-\beta)^2}{\epsilon^2(a/\nu)^2(1-t/T)}\right) + O\left(\epsilon \cdot \max(\epsilon, |\rho|)\right), \tag{32}$$

where $\Gamma(\theta, z) = \int_z^\infty e^{-x}x^{\theta-1}dx$ and $\Gamma(\theta) = \Gamma(\theta, 0)$ are the upper incomplete gamma and gamma functions, respectively. Note that the first term on the right hand side of (32) converges faster than any polynomial order of ϵ. Therefore, by the formulas (30), (31), and (32), the price for the put option without arbitrage can be approximated by $\bar{\varphi}_p(t, f, a)$ as follows:

$$V_p(t, f, a) = \bar{\varphi}_p(t, f, a) + O(\epsilon \cdot \max(\epsilon, |\rho|)). \tag{33}$$

We can summarize the above results in the following theorem.

Theorem 2. *Under* Assumption 1*, the prices for the call and put options without arbitrage, i.e., $V_c(t, f, a)$ and $V_p(t, f, a)$, can be expanded in* (28) *and* (33)*, respectively.*

Note that the approximate formulas in (27) and (31) involve the noncentral chi-square distribution function, which can be evaluated very quickly and thus be applied in real time transactions. There exist a lot of efficient algorithms to implement the noncentral chi-square distribution function. One can refer to Dyrting (2004) and Larguinho et al. (2013), which discuss the tradeoff between speed and accuracy for different approaches. They both provide a comprehensive literature review and detailed numerical tests for the performance of various approaches. For convenience and completeness, we also present and test some methods in detail in Appendix D. In the numerical experiments, for the purpose of only illustrating the validity of our formulas, we directly use the library function '*ncx2cdf*' in *Matlab*.

4. Numerical Experiments

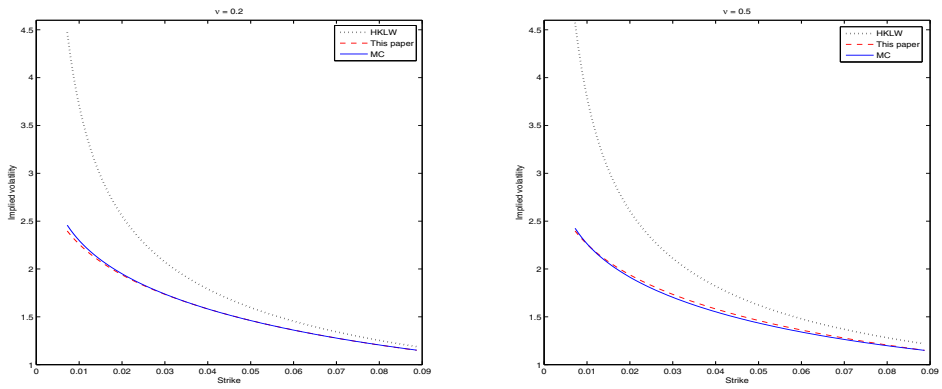
This section verifies the accuracy and effectiveness of our approximate formulas (27) and (31) for the call and put options without arbitrage. For the formula (27), implied volatilities rather than option prices are examined since the implied volatility is modeled directly in practice. The

approximate formula (31) for the put option price is directly examined. The codes for all numerical experiments are written in Matlab R2013a, and implemented on a PC desktop with Intel(R) Core(TM)2 Quad CPU Q9400@2.66GHZ.

We compare the performance of implied volatilities computed from three approaches: the closed-form formula provided by Hagan et al. (2002), our approximate formula (27) for the call option price, and the Monte Carlo simulation. First, the implied volatilities of Hagan et al. (2002) are directly computed from their formula. Second, we first calculate the call option prices using (27), and then obtain the implied volatilities by numerically inverting the option prices. The calculation of option prices through (27) and implied volatilities are accomplished by two library functions '*ncx2cdf*' and '*blsimpv*' in Maltab, respectively. Such computation only takes several milliseconds on average for one trial. Third, the Monte Carlo simulations via Euler discretization produce the call option prices without arbitrage based on (4). Then we also numerically inverted the option price to the implied volatility. To monitor the absorbing boundary at zero, the number of simulated samples is $1,000,000$ and the number of time steps is $25,200$ per year, which apply in this whole section.

The performance of these approaches is examined in three aspects. We compute the implied volatilities with different levels of strike, vol-of-vol, and correlation. We first plot implied volatilities generated by three methods against strike prices in Figures 1-4, which correspond to different vol-of-vol, correlation, initial volatility, and beta (i.e., the parameter β in (1)), respectively. The blue solid line, denoted by "MC", is the implied volatilities generated by the Monte Carlo simulations. The black dotted line, denoted by "HKLW", is plotted from the implied volatility formula of Hagan et al. (2002). The red dashed line, denoted by "This paper", is generated by the formula (27). Second, we plot errors of the formulas (27) and Hagan et al. (2002) relative to the Monte Carlo simulations (benchmark) in Figure 5 when the vol-of-vol varies between 0.1 and 4 provided four maturities "1/12, 1/4, 1/2, 1". The black dotted and red dashed lines are plotted from Hagan et al. (2002) and the formula (27), respectively. There are 40 points in each line. The relative error is defined by "Relative error $= \frac{|\text{Method}-\text{MC}|}{\text{MC}}$", where "Method" is either "HKLW" or "This paper". Third, we plot the relative errors in Figure 6 as the correlation changes from "-0.1" to "-0.9". The rest of the other setting is similar to that in Figure 5.

Figure 1: Implied Volatilities versus Strikes under Different Levels of Volatility of Volatility



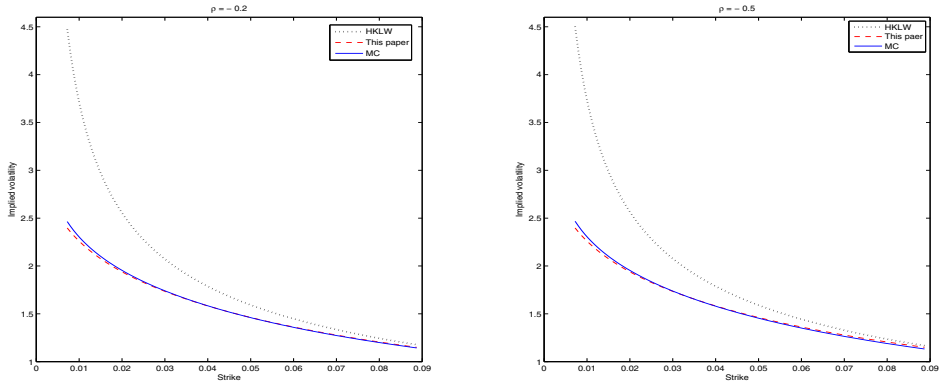
Notes: The black dotted, blue solid, and red dashed lines plot the implied volatilities generated by Hagan et al. (2002)'s formula, the Monte Carlo simulation, and the formula (27), respectively. In the left panel, the volatility of volatility is $\nu = 0.2$, while in the right panel it is $\nu = 0.5$. The rest of the parameters are $f = 0.05, T = 1, a = 0.1, \beta = 0.1, \rho = 0$.

Figures 1-4 plot implied volatilities against strikes under different combinations of other parameters. The pictures on the left and right panels in Figure 1 correspond to relatively small and large vol-of-vol, respectively, while the other parameters values are the same for both pictures. For the two levels of vol-of-vol, the implied volatilities generated by the formula (27) are quite close to the benchmark (the Monte Carlo simulation), whereas the implied volatility from Hagan et al. (2002) deviates from the benchmark as the strike price decreases. The deviation becomes substantially large for small strikes. Similar patterns appear in Figure 2, in which the correlations for pictures on the left and right are relatively small and large, respectively.

In Figure 3, the pictures on the left and right panels have relatively small and large initial volatilities, respectively. Implied volatilities from Hagan et al. (2002) are inconsistent with the benchmark. For the formula (27), its performance may also become inconsistent relative to the benchmark as the strikes become extremely small. However, the overall performance of our formula (27) is better than that of Hagan et al. (2002).

For the picture on the left panel of Figure 4, our results almost overlap with the benchmark when the beta is small ($\beta = 0.2$). In contrast, the distance between Hagan et al. (2002)'s implied

Figure 2: Implied Volatilities versus Strikes under Different Degrees of Correlation

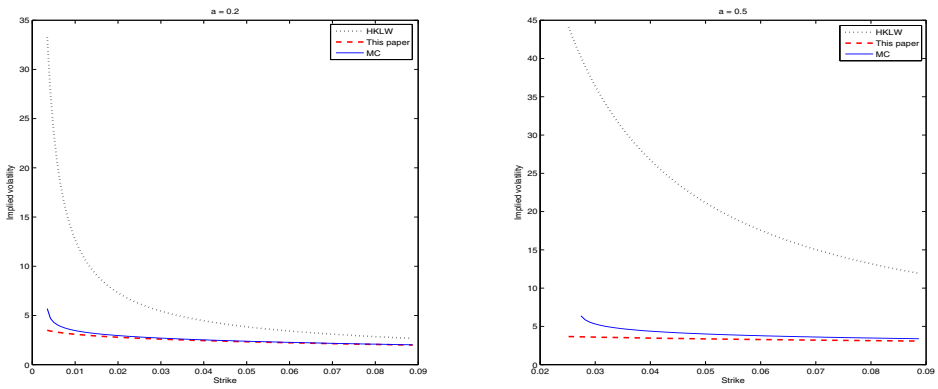


Notes: The black dotted, blue solid, and red dashed lines plot the implied volatilities generated by Hagan et al. (2002)'s formula, the Monte Carlo simulation, and the formula (27), respectively. In the left panel, the correlation is $\rho = 0.2$, while in the right panel it is $\rho = 0.5$. The rest of the parameters are $f = 0.05$, $T = 1$, $a = 0.1, \beta = 0.1, \nu = 0.1$.

volatilities and the benchmark cannot be neglected. For the picture on the right panel of Figure 4 with a larger beta ($\beta = 0.5$), the implied volatilities generated by all three methods coincide with each other for most of the strikes. However, for extremely small strikes, the implied volatilities from both the formula (27) and Hagan et al. (2002) are almost the same, and a little larger than the benchmark.

Figure 5 demonstrates the impact of the total vol-of-vol ($\epsilon = \nu\sqrt{T}$) on the validity of the formula (27). For a given maturity ($T = 1/12, 1/4, 1/2, 1$), each picture plots the relative errors produced by the formula (27) and Hagan et al. (2002) as the vol-of-vol changes from 0.1 to 4. The magnitude of the total vol-of-vol instead of the vol-of-vol determines the region valid for the approximation. In detail, the relative errors of the formula (27) are less than 1% for the following cases: (i) $T = 1/12$ and $\epsilon \leq 0.37$, (ii) $T = 1/4$ and $\epsilon \leq 1$, (iii) $T = 1/2$ and $\epsilon \leq 0.35$, (ii) $T = 1$ and $\epsilon \leq 0.2$. If we increase the tolerance level to 5%, then the formula (27) is valid for: (i) $T = 1/12$ and $\epsilon \leq 1.13$, (ii) $T = 1/4$ and $\epsilon \leq 1.55$, (iii) $T = 1/2$ and $\epsilon \leq 0.98$, (iv) $T = 1$ and $\epsilon \leq 0.6$. Moreover, Figure 5 also shows that the smaller the maturity, the larger the total vol-of-vol valid for the approximation. In contrast, the relative errors of Hagan et al. (2002) are larger

Figure 3: Implied Volatilities versus Strikes under Different Initial Volatilities



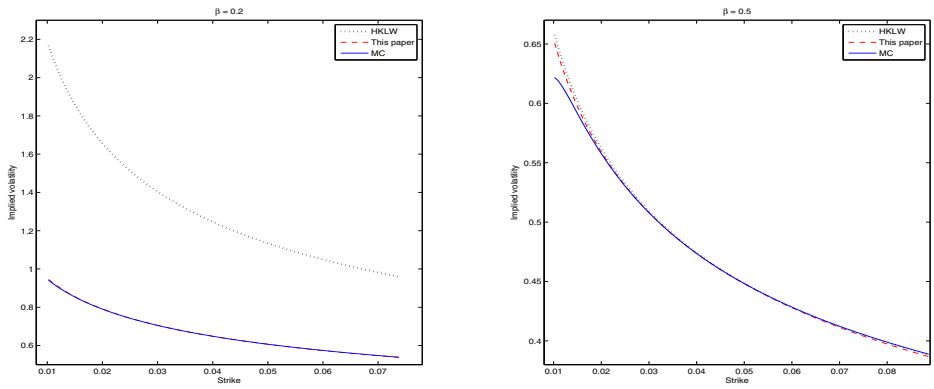
Notes: The black dotted, blue solid, and red dashed lines plot the implied volatilities generated by Hagan et al. (2002)'s formula, the Monte Carlo simulation, and the formula (27), respectively. In the left panel, the initial volatility is $a = 0.2$, while in the right panel it is $a = 0.5$. The rest of the parameters are $f = 0.05$, $T = 1, \beta = 0.1, \rho = 0, \nu = 0.1$.

than 5% for $T = 1$ and 1% for $T = 1/2$. The last three pictures indicate that our formula (27) always outperform Hagan et al. (2002)'s formula. It is worth noting that as the vol-of-vol becomes extremely large, both methods produce biased results for $T = 1/4, 1/2, 1$.

Figure 6 studies how the correlation ρ affects the accuracy of the approximation. Similarly, for a given maturity, each picture plots the relative errors of the formula (27) and Hagan et al. (2002) as the correlation varies from -0.9 to -0.1. For all combinations of the correlation and the maturity, the relative errors of the formula (27) are less than 1%. However, the closed-form formula of Hagan et al. (2002) leads to more significant relative errors or even biased results. Specifically, the relative errors of Hagan et al. (2002)'s formula are (i) less than 1% for $T = 1/12, 1/4$; (ii) about 2-5% for $T = 1/2$; (iii) 9-10% for $T = 1$. For short maturities ($T = 1/12, 1/4$), both methods perform very well. Nonetheless, for relatively long maturities ($T = 1/2, 1$), the performance of the formula (27) is much better than that of Hagan et al. (2002) for different correlations in this case.

To test the validity of the formula (31), we use it to compute the put option price and compare with the results from Hagan et al. (2002) and the Monte Carlo simulation (the benchmark). The Monte Carlo simulation of the put option price without arbitrage is based on the formula (5).

Figure 4: Implied Volatilities versus Strikes under Different Betas



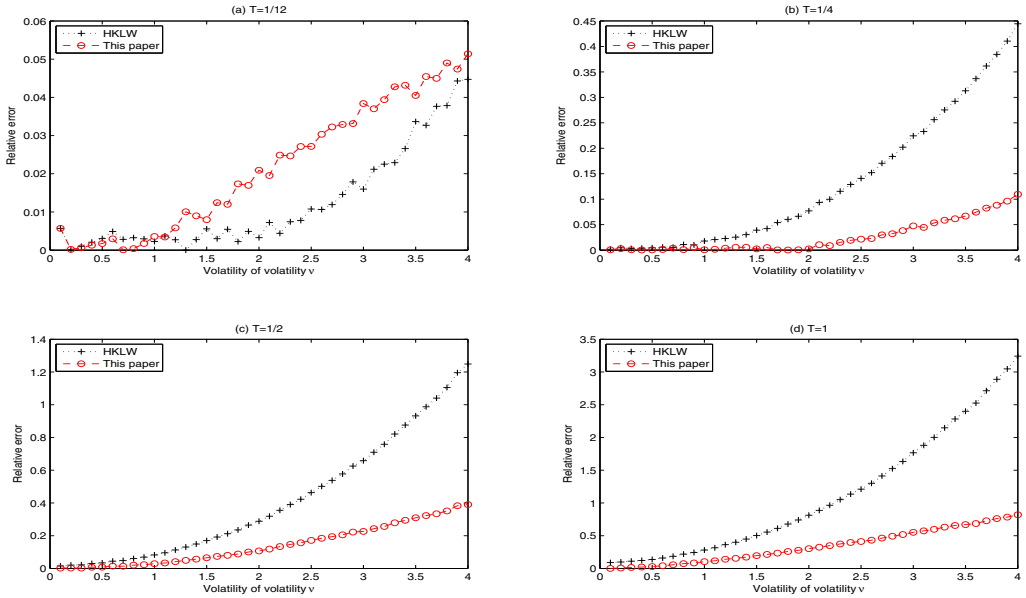
Notes: The black dotted, blue solid, and red dashed lines plot the implied volatilities generated by Hagan et al. (2002)'s formula, the Monte Carlo simulation, and the formula (27), respectively. In the left panel, the index of the CEV component beta is $\beta = 0.2$, while in the right panel it is $\beta = 0.5$. The rest of the parameters are $f = 0.05$, $T = 1$, $a = 0.1$, $\rho = 0$, $\nu = 0.1$.

Hence the probability that the forward price hits zero has been tracked. We examine the relative errors generated by different approaches against the vol-of-vol, the correlation, and the strike in Figures 7, 8, and 9, respectively.

For each given maturity, Figure 7 plots the relative errors of the put option prices, which is defined similar to that of implied volatilities, against the vol-of-vol in each picture. Given that the relative error is less than 1%, the formula (31) is valid when ϵ is smaller than 0.46, 0.75, 0.35, 0.1 for the four maturities $T = 1/12,\ 1/4,\ 1/2,\ 1$, respectively. If the relative error is less than 5%, then the formula (31) is valid when ϵ is smaller than 1.15, 1.3, 0.71, 0.5, respectively. For $T = 1$ and $1/2$, the relative errors of the put option prices from Hagan et al. (2002) are always larger than 5% and 1%, respectively. Though both methods may be biased for $T = 1/4,\ 1/2,\ 1$ for particularly larger vol-of-vol, all four pictures suggest that the performance of the formula (31) is preferable to that of Hagan et al. (2002).

Figure 8 illustrates how the performance of each method is affected when the correlation varies between -0.9 and -0.1. The relative error from our formula (31) is not sensitive to the variation of the correlation because they are all less than 2% for all four maturities. This observation also

Figure 5: Relative Errors of Implied Volatilities versus Volatilities of Volatility under Different Maturities

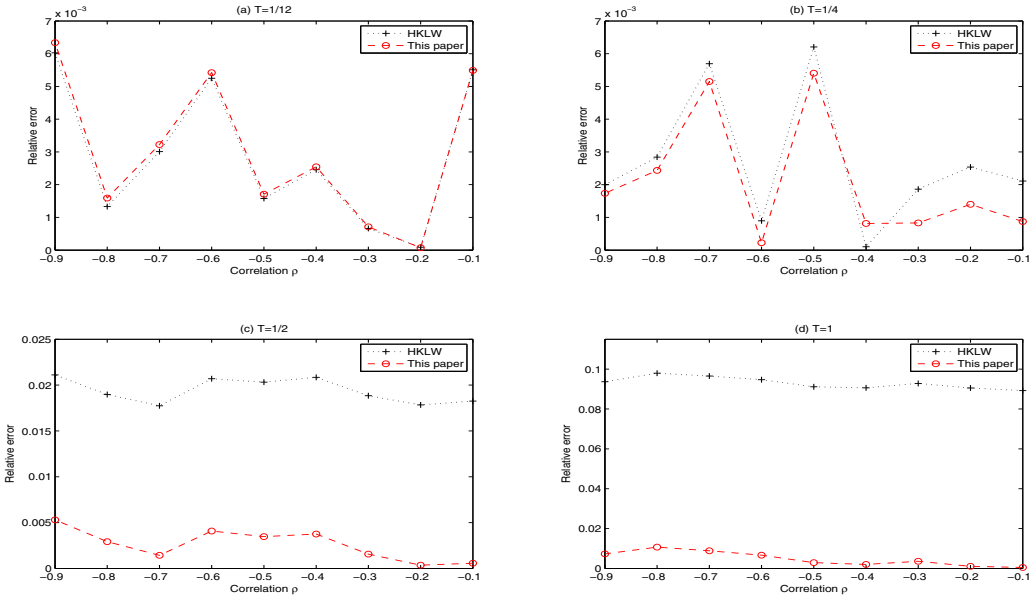


Notes: "HKLW" is the relative error of Hagan et al. (2002), which is indicated by the black dotted line with pluses. "This paper" is the relative error of the formula (27), which is the red dashed line with circles. The relative error is obtained by "Relative error $= \frac{|\text{Method}-\text{MC}|}{\text{MC}}$", where "Method" is either "HKLW" or "This paper". The parameters values are $f = 0.05$, $a = 0.1$, $\beta = 0.1$, $\rho = -0.2$, and $K = 0.05$.

numerically confirms that the effect of the correlation is not significant. On the contrary, Hagan et al. (2002) has unfavorable relative errors (8%-9%) for $T = 1$. Figure 9 plots the relative errors against the strike prices in each picture. Compared with the benchmark, the performance of the formula (31) is much better that of Hagan et al. (2002). In contrast, for extremely small strikes, Hagan et al. (2002) leads to significantly biased results for all four maturities.

In addition, Table 1 presents the performance of arbitrage-free option prices based on formulas (27) and (31) for extremely large maturities, ranging from 1 year to 25 years. The errors of call prices from (27) relative to the Monte Carlo simulation are smaller than 1% for all maturities. The relative errors of put prices from (31) are all less than 1.4%. Therefore, at least for this group of

Figure 6: Relative Errors of Implied Volatilities versus Correlations under Different Maturities

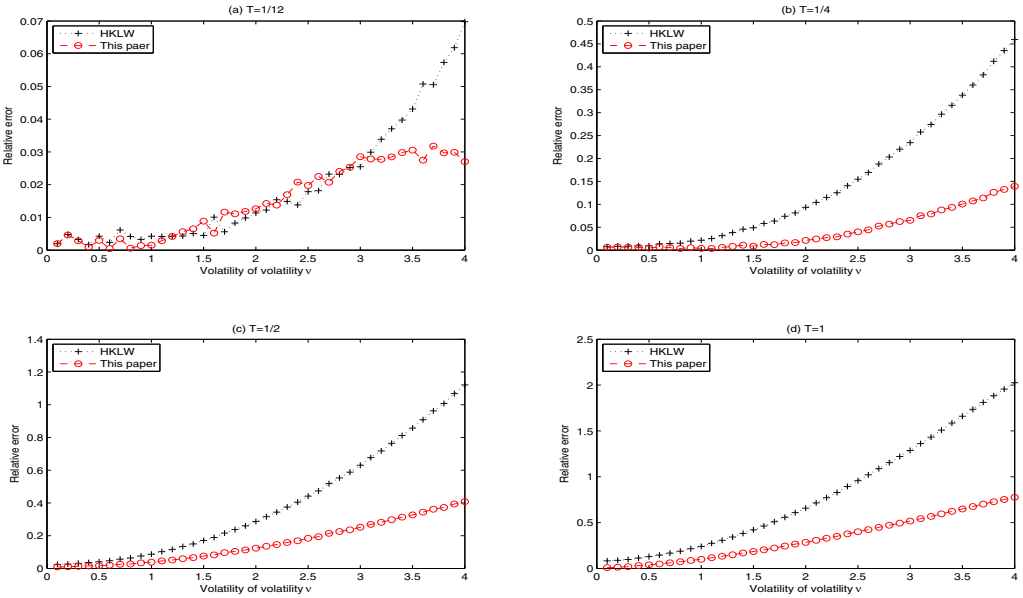


Notes: "HKLW" is the relative error of Hagan et al. (2002), which is indicated by the black dotted line with pluses. "This paper" is the relative error of the formula (27), which is the red dashed line with circles. The relative error is obtained by "Relative error $= \frac{|\text{Method}-\text{MC}|}{\text{MC}}$", where "Method" is either "HKLW" or "This paper". The parameters values are $f = 0.05$, $a = 0.1$, $\beta = 0.1$, $\nu = 0.1$, and $K = 0.05$.

parameters, the formulas (27) and (31) are valid for very long maturities, which can be attributed to that the accuracy of our approximation is jointly determined by the product of the vol-of-vol and the maturity, i.e., the total vol-of-vol. In this case, the total vol-of-vol ranges from 0.1 to 0.5.

In sum, the overall performance of our approximate formulas is comparable with the benchmark (the Monte Carlo simulation) for small total vol-of-vol. First, for almost all levels of strike, vol-of-vol, and correlation we have tested, the performance of our formulas (27) and (31) are better than that of Hagan et al. (2002). Second, for almost all strikes under different other parameter combinations, our formulas (27) and (31) produce results which are comparable to the benchmark. While Hagan et al. (2002) often leads to unsatisfactory results especially for small strikes. Third,

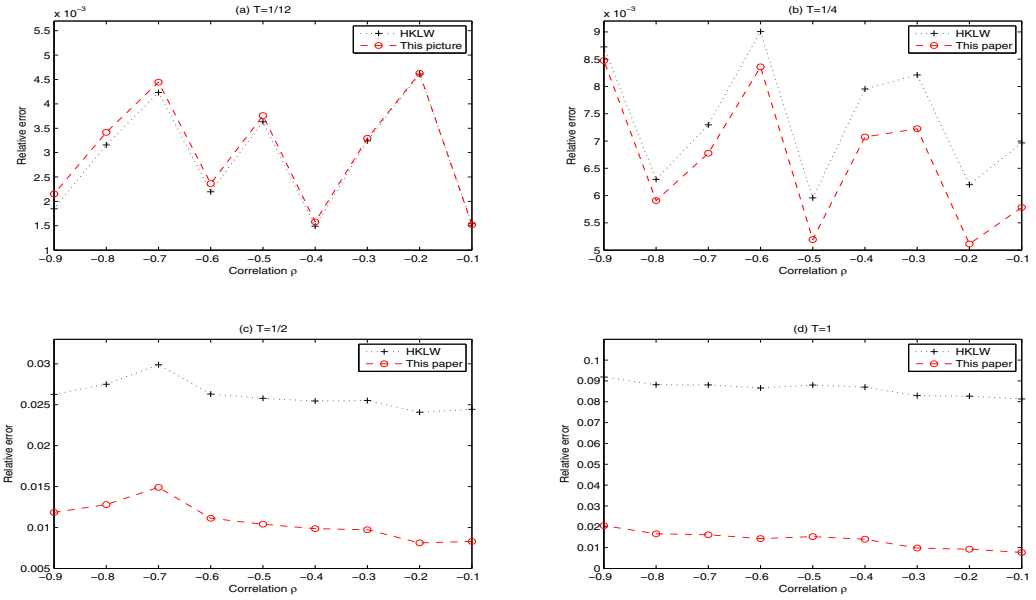
Figure 7: Relative Errors of Put Option Prices versus Volatilities of Volatility under Different Maturities



Notes: "HKLW" is the relative error of Hagan et al. (2002), which is indicated by the black dotted line with pluses. "This paper" is the relative error of the formula (31), which is the red dashed line with circles. The relative error is obtained by "Relative error $= \frac{|\text{Method}-\text{MC}|}{\text{MC}}$", where "Method" is either "HKLW" or "This paper". The parameters values are $f = 0.05$, $a = 0.1$, $\beta = 0.1$, $\rho = -0.2$, and $K = 0.05$.

the vol-of-vol and the maturity, i.e., total vol-of-vol, jointly determine the valid region for the approximation. Specifically, comparing with the benchmark, the performance of our formulas is quite good when the total vol-of-vol is small. However, both our and Hagan et al. (2002)'s formulas generate in biased results for large total vol-of-vol. Finally, the formulas (27) and (31) perform quite well across different correlations, which justifies the fact that the high order terms only with respect to the correlation in the expansion are zero. When the maturity is one year, Hagan et al. (2002) may lead to unfavorable results for different correlations.

Figure 8: Relative Errors of Put Option Prices versus Correlations under Different Maturities

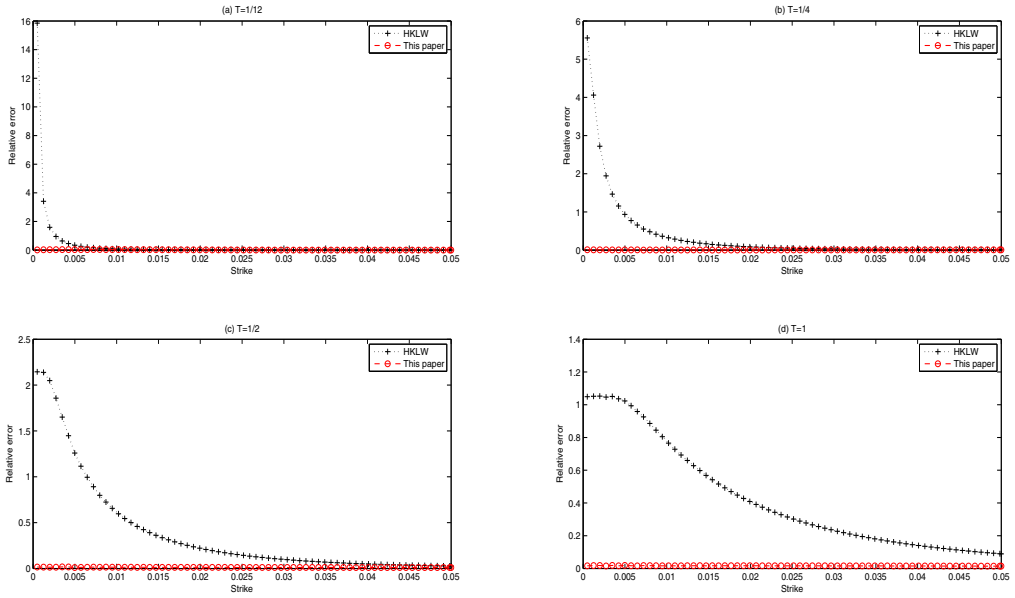


Notes: "HKLW" is the relative error of Hagan et al. (2002), which is indicated by the black dotted line with pluses. "This paper" is the relative error of the formula (31), which is the red dashed line with circles. The relative error is obtained by "Relative error $= \frac{|\text{Method}-\text{MC}|}{\text{MC}}$", where "Method" is either "HKLW" or "This paper". The parameters values are $f = 0.05$, $a = 0.1$, $\beta = 0.1$, $\nu = 0.1$, and $K = 0.05$.

5. Conclusions

The SABR model is widely used in practice and has become the benchmark model in interest rate and foreign exchange markets. However, the arbitrage-free vanilla option pricing formulas under the SABR model remain unknown. In this paper, we derive explicit formulas to approximate the vanilla option prices under the SABR model. Our formulas have several appealing features. First, they are arbitrage-free because we have allowed for an absorbing boundary at zero. Second, they are easy to implement. We can use either the library functions of most numerical software packages or some efficient algorithms to evaluate them quickly, though they involve the noncentral chi-square distribution function. Third, the analytical formulas lead to fast and effective numer-

Figure 9: Relative Errors of Put Option Prices versus Strikes under Different Maturities



Notes: "HKLW" is the relative error of Hagan et al. (2002), which is indicated by the black dotted line with pluses. "This paper" is the relative error of the formula (31), which is the red dashed line with circles. The relative error is obtained by "Relative error $= \frac{|\text{Method}-\text{MC}|}{\text{MC}}$", where "Method" is either "HKLW" or "This paper". The parameters values are $f = 0.05$, $a = 0.1$, $\beta = 0.1$, $\rho = -0.2$, and $\nu = 0.2$.

ical results, and thus can be applied to computing implied volatilities for real time transactions. Finally, the extensive numerical experiments show the effectiveness of our approximate formulas especially when the total vol-of-vol is small.

Acknowledgements

We are grateful to the conference participants at 9th World Congress of the Bachelier Finance Society for their comments. This work was supported in part by the National Science Foundation of China [Grants 71231008, 71401103, 71421002, 71501196 and 71721001], the Research Grant Council of Hong Kong SAR Government [Project ID: CUHK14201114], the Natural Science Foundation of Guangdong Province of China [Grant 2014A030312003], the Fundamental

Table 1: Relative Errors under Long Maturities

T	1	2	3	4	5	10	15	20	25
R.E. of (27)	0.31%	0.07%	0.16%	0.05%	0.27%	0.01%	0.46%	0.94%	0.92%
R.E. of (31)	0.96%	1.00%	1.11%	1.08%	1.13%	1.28%	1.25%	1.40%	1.35%

Notes: The first row is the values of different maturities. The second row is the relative errors of call option prices from the formula (27). The third row is the relative errors of put option prices from the formula (31). Other parameters values are $\rho = -0.2$, $f = 0.05$, $a = 0.1$, $\beta = 0.1$, $\nu = 0.1$, and $K = 0.05$.

Research Funds for the Central Universities of China [Grants 010414370101 and 14wkpy63] and research grant from Advanced Research Institute of Finance, Sun Yat-sen University, respectively.

Appendix A. Proof of Theorem 1

First of all, the stochastic process (F, A) defined in (1) admits a solution, see Theorem 3.1 in Hobson (2010). Let $S_n = \inf\{t > 0 : F_t^2 + A_t^2 \geq n^2\}$ and $\tau_n = \inf\{s > t : F_s \leq \frac{1}{n}\}$. Apparently, as $n \to +\infty$, the limit of τ_n is τ_t (cf. (3)). Applying Itô's formula to $\varphi(t, F_t, A_t)$ and recalling PDE (6) for $\varphi(t, f, a)$, we have

$$\varphi(T \wedge \tau_n \wedge S_n, F_{T\wedge\tau_n\wedge S_n}, A_{T\wedge\tau_n\wedge S_n}) = \varphi(t, F_t, A_t) + \int_t^{T\wedge\tau_n\wedge S_n} \frac{\partial\varphi}{\partial f} dF_{t'} + \int_t^{T\wedge\tau_n\wedge S_n} \frac{\partial\varphi}{\partial a} dA_{t'}.$$

Taking expectation on both sides, the resulting stochastic integrals have zero expectations. Thus we have

$$\varphi(t, f, a) = \mathbb{E}[\varphi(T \wedge \tau_n \wedge S_n, F_{T\wedge\tau_n\wedge S_n}, A_{T\wedge\tau_n\wedge S_n})|F_t = f, A_t = a].$$

Note that $|\varphi(t, f, a)| \prec C(1 + |f|)$, where C is a positive constant. Moreover, by Proposition 5.1 of Andersen and Piterbarg (2007), the process F has finite moments, e.g., $\mathbb{E}[|F_t|] < +\infty$. Recall the boundary and terminal conditions for $\varphi(t, f, a)$ in (7). As $n \to +\infty$, by the dominated convergence theorem, we have

$$\varphi(t, f, a) = \mathbb{E}[h(F_T)\mathbf{1}_{\{\tau_t > T\}}|F_t = f, A_t = a].$$

Finally, the uniqueness of the solution is a direct consequence of the stochastic representation above by taking $h(f) \equiv 0$. The proof completes.

Appendix B. Calculation of the Formula (24)

By the definitions of $\Lambda(\tau, u, \xi)$ in (23) and $q(\xi; \kappa, y)$ in (25), the following relationship holds

$$\Lambda(\tau, x, \xi) = \frac{2\xi}{\tau} \cdot q(\frac{x^2}{\tau}; 2\theta + 2, \frac{\xi^2}{\tau}) = \frac{x^{2\theta}}{\xi^{2\theta}} \frac{2\xi}{\tau} \cdot q(\frac{\xi^2}{\tau}; 2\theta + 2, \frac{x^2}{\tau}). \tag{B.1}$$

Let $A_1 = y^{2\theta} \int_{k/y}^{\infty} \Lambda(\tau, x, \xi)\xi^{2\theta} d\xi$ and $A_2 = k^{2\theta} \int_{k/y}^{\infty} \Lambda(\tau, x, \xi) d\xi$. Then,

$$C^{(0)}(\tau, x, y) = \int_0^{+\infty} \Lambda(\tau, x, \xi)((\xi y)^{2\theta} - k^{2\theta})^+ d\xi = A_1 - A_2, \tag{B.2}$$

Using the equalitise in (B.1), we can rewrite A_1 and A_2 as follows:

$$A_1 = x^{2\theta} y^{2\theta} \int_{\frac{k^2}{\tau y^2}}^{\infty} q(z; 2\theta + 2, \frac{x^2}{\tau}) dz, \quad A_2 = k^{2\theta} \int_{\frac{k^2}{\tau y^2}}^{\infty} q(\frac{x^2}{\tau}; 2\theta + 2, z) dz.$$

Recalling the noncentral chi-square distribution function $Q(x; \kappa, y)$ defined in (25), we have that

$$A_1 = (xy)^{2\theta}(1 - Q(\frac{k^2}{\tau y^2}; 2\theta + 2, \frac{x^2}{\tau})). \tag{B.3}$$

Note the following equality holds (see, e.g., Schroder, 1989; Lesniewski, 2009),

$$\int_x^{+\infty} q(z; \kappa, y) dz + \int_y^{+\infty} q(x; \kappa + 2, z) dz = 1.$$

Then we have an explicit formula for A_2 as follows:

$$A_2 = k^{2\theta} Q(\frac{x^2}{\tau}; 2\theta, \frac{k^2}{\tau y^2}). \tag{B.4}$$

By the formulas (B.3) for A_1, (B.4) for A_2, and (B.2), then we have $C^{(0)}(\tau, x, y)$ in (24).

Appendix C. Derivation of the Formula (31)

Mimicking the derivation of the formula (27), there are four steps to derive the approximate formula in (31) for the put. We first define a function $P_1(\tau, f, g)$ for the put option price after the *rescaling* (8), whereby $P_1(\tau, f, g) := \varphi_p(t, f, a) \equiv \varphi_p(T(1 - \tau), f, \nu g)$ and $P_1(\tau, f, g)$ satisfies the PDE (10) with the same boundary condition and a different initial condition $P_1(0, f, g) = (K - f)^+$.

Second, after performing the Lamperti transformation (11), define a new function $P_2(\tau, x, y) := P_1(\tau, f, g) \equiv P_1(\tau, (\epsilon(1 - \beta)xy)^{1/(1-\beta)}, y)$. Moreover, $P_2(\tau, x, y)$ satisfies the PDE in (13) with the

same boundary condition and a different initial condition $P_2(0,x,y) = (k-(\epsilon(1-\beta)xy)^{2\theta})^+$. Third, the *homogenization* (18) allows us to obtain an initial value without perturbation parameters. Thus, the new function is $P(\tau,x,y) := (\epsilon(1-\beta))^{-2\theta}P_2(\tau,x,y) \equiv (\epsilon(1-\beta))^{-2\theta}P_1(\tau,(\epsilon(1-\beta)xy)^{2\theta},y)$. Similarly, $P(\tau,x,y)$ satisfies the PDE in (19) with the same boundary condition and a different initial condition $P(0,x,y) = (k-(xy)^{2\theta})^+$.

Finally, we can expand $P(\tau,x,y)$ first w.r.t. ϵ and then w.r.t. ρ. Then, the leading order term $P_\rho^{(0)}(\tau,x,y)$ satisfies the PDE in (21) with the different initial condition $P_\rho^{(0)}(0,x,y) = (k-(xy)^{2\theta})^+$ and the same boundary condition. Moreover, the first order term and higher order terms with respect to the correlation are also zero. Using arguments similar to those in Section 3.2, we have

$$P(\tau,x,y) = P_\rho^{(0)}(\tau,x,y) + O(\epsilon\cdot\max(\epsilon,|\rho|)),$$
$$P^{(0)}(\tau,x,y) = k^{2\theta}(1-Q(\frac{x^2}{\tau};2\theta,\frac{k^2}{\tau y^2})) - (xy)^{2\theta}Q(\frac{k^2}{\tau y^2};2\theta+2,\frac{x^2}{\tau}).$$

Reversing the above procedure, we can show that the approximate formula for the price of the down-and-out put option under the original coordinates is given by $\bar{\varphi}_p(t,f,a)$ in (31).

Appendix D. A Digestion on the Computation of the Noncentral Chi-Square Distribution

The computation of the noncentral chi-square distribution is well documented in the literature since this distribution is widely used in finance and statistics. Two major approaches, the gamma series method and analytical approximations, have been developed to compute the noncentral chi-square distribution. Other typical methods include the asymptotic expansion by Temme (1993) and the Bessel series method by Dyrting (2004). This distribution function can be easily implemented using softwares such as *Mathlab* and *Mathematica*. One can refer to Dyrting (2004) and Larguinho et al. (2013) for a comprehensive literature review and detailed numerical tests on the performance of various methods. We then provide details of the gamma series method and analytical approximations.

Appendix D.1. The Gamma Series Method

The cumulative and complementary distribution functions of the noncentral chi-square distribution, i.e., $Q(x;\kappa,y)$ and $Q^c(x;\kappa,y)$, can be expressed as series of incomplete gamma functions

(see, e.g., Ding, 1992; Dyrting, 2004; Larguinho et al., 2013; Schroder, 1989)

$$Q(x;\kappa,y)=\sum_{i=0}^{\infty}\frac{(y/2)^i e^{-y/2}}{i!}\frac{\gamma(\kappa/2+i,x/2)}{\Gamma(\kappa/2+i)},\quad Q^c(x;\kappa,y)=\sum_{i=0}^{\infty}\frac{(y/2)^i e^{-y/2}}{i!}\frac{\Gamma(\kappa/2+i,x/2)}{\Gamma(\kappa/2+i)},\tag{D.1}$$

where $\Gamma(\kappa)=\int_0^{\infty}\xi^{\kappa-1}e^{-\xi}d\xi$, $\gamma(\kappa,x)=\int_0^{x}\xi^{\kappa-1}e^{-\xi}d\xi$, and $\Gamma(\kappa,x)=\Gamma(\kappa)-\gamma(\kappa,x)$ are gamma, lower incomplete gamma, and upper incomplete gamma functions, respectively. Fraser et al. (1998) and Larguinho et al. (2013) have applied the gamma series method for computing exact probabilities directly. Carr and Linetsky (2006) also use it to calculate option prices in a jump to default extended CEV model.

There are two approaches to evaluate the gamma series more efficiently. First, Schroder (1989) and Ding (1992) reduce the gamma series into a double series of the gamma function, whereby the gamma function can be evaluated through elementary functions, see, e.g., Press et al. (1992). Second, Knüsel and Bablok (1996) and Benton and Krishnamoorthy (2003) (advocated by Dyrting, 2004; Larguinho et al., 2013 respectively) evaluate the summation by determining the number of terms in the summation for a given level of error tolerance. Moreover, two consecutive terms in the gamma series have a recurrence relation, which allows us to generate each term using the previous one and determine the precise number of terms.

Appendix D.2. Analytical Approximations

For large argument x and noncentrality y (perhaps either $x>1000$ or $y>1000$, see, e.g., Benton and Krishnamoorthy, 2003; Dyrting, 2004), the analytical approximation is a method of choice to compute the distribution function $Q(x;\kappa,y)$. Specifically, the noncentral chi-square distribution can be approximated by a standard normal normal distribution, that is,

$$Q(x/2;\kappa/2,y/2)\approx\Phi(z),$$

where $\Phi(\cdot)$ is the standard normal distribution function. A typical approximation is provided by Sankaran (1963), where

$$z=-\frac{1-hp(1-h+(2-h)mp/2)-(x/(\kappa+y))^h}{h\sqrt{2p(1+mp)}},$$

where $h=1-\frac{2}{3}\frac{(\kappa+y)(\kappa+3y)}{(\kappa+2y)^2}$, $p=\frac{1}{2}\frac{\kappa+2y}{(\kappa+y)^2}$ and $m=(h-1)(1-3h)$. Moreover, this approximation is suggested by Schroder (1989).

Appendix D.3. Numerical Tests

We test the speed and accuracy of the gamma series method and the analytical approximation by Sankaran (1963). The noncentral chi-square distribution function $Q(x;\kappa,y)$ is evaluated by the gamma series method (D.1), the analytical approximation of Sankaran (1963) and the library function '*ncx2cdf*' of *Matlab*, in a cube $[0.0001, 300.0001]\times[1,5]\times[0.0001, 300.0001]$, respectively. The grid selected covers the parameters values used in this paper. In each dimension, the length of the step is one. Thus, the function is evaluated in 450,000 points. Since the time for computing the function once is very small, the time in the table below indicates the total time for computing 450,000 function values. The results are summarized in the table below.

Table D.2: Testing the Algorithms to Compute the Noncentral Chi-square Distribution Function

	Matlab	Gamma Series (D.1)	Sankaran (1963) A	Sankaran (1963) B
Error	NA	1.11E-16	0.010973	0
Time	118.19	51.74	18.97	NA

Notes: The row "Error" denotes the maximum absolute error between the values computed from the corresponding method and *Matlab*. Specifically, "Sankaran (1963) A" measures the error in the whole cube, while "Sankaran (1963) B" measures the error for either $x \in [100.0001, 300.0001]$ or $y \in [100.0001, 300.0001]$. The row "Time" denotes the total running time for the corresponding method, which is measured in seconds for total 450,000 function values. "NA" means "not applicable" for that case.

Table D.2 indicates that a combination of the gamma series (D.1) and the analytical approximation Sankaran (1963) can evaluate the noncentral chi-square function more efficiently without sacrificing accuracy. Specifically, we evaluate the function for small argument x or small noncentrality y using the gamma series method. Once either x or y is larger than 100, we switch to the analytical approximation Sankaran (1963).

References

References

Andersen, L., Piterbarg, V., 2007. Moment explosions in stochastic volatility models. Finance Stoch. 11(1), 29–50.

Balland, P., Tran, Q., 2013. SABR goes normal. Risk 26(6), 72–77.

Benton, D., Krishnamoorthy, K., 2003. Computing discrete mixtures of continuous distributions: noncentral chisquare, noncentral t and the distribution of the square of the sample multiple correlation coefficient. Comput. Statist. Data Anal. 43, 249–267.

Borodin,N., Salminen, P., 2002. Handbook of Brownian Motion-Facts and Formulae, second ed. Birkhäuser, Basel.

Brigo, D., Mercurio, F., 2006. Interest Rate Models: Theory and Practice with Smile, Inflation and Credit, second ed. Springer Verlag, New York.

Carr, P., Linetsky, V., 2006. A jump to default extended CEV model: an application of Bessel processes. Finance Stoch. 10(3), 303-330.

Davydov, D., Linetsky, V., 2001. Pricing and hedging path-dependent options under the CEV process. Management Sci. 47(7), 949–965.

Delbaen, F., Shirakawa, H., 2002. A note on option pricing for the constant elasticity of variance model. Asia-Pac. Financ. Markets 9(2), 95–99.

Ding, C., 1992. Algorithm AS 275: computing the non-central χ^2 distribution function. Appl. Statist. 41, 478–482.

Doust, P., 2012. No-arbitrage SABR. J. Comput. Finance 110(3), 3–31.

Dyrting, S., 2004. Evaluating the noncentral chi-square distribution for the Cox-Ingersoll-Ross process. Computat. Econ. 24, 35–50.

Fraser, D., Wu, J., Wong, A., 1998. An approximation for the noncentral chi-squared distribution. Commun. Statist.-Simul. Comput. 27, 275–287.

Fouque, J.-P., Papanicolaou, G., Sircar, R., 2000. Derivatives in Financial Markets with Stochastic Volatility. Cambridge University Press, Cambridge, UK.

Fouque, J.-P., Papanicolaou, G., Sircar, R., Solna, K., 2003. Multiscale stochastic volatility asymptotics. Multiscale Model. Simul., 2, 22–42.

Gulisashvili, A., Horvath, B., Jacquier, A., 2015. Mass at zero for small-strike implied volatility expansion in the SABR model. Available at: arXiv:1502.03254.

Hagan, P.S., Kumar, D., Lesniewski, A.S., Woodward, D.E., 2002. Managing smile risk. Wilmott Magazine 1, 84–108.

Hagan, P.S., Kumar, D., Lesniewski, A.S., Woodward, D.E., 2014. Arbitrage-free SABR. Wilmott Magazine 69, 60–75.

Hagan, P.S., Lesniewski, A.S., Woodward, D.E., 2015. Probability distribution in the SABR model of stochastic volatility, in: Friz, P., Gatheral, J., Gulisashvili, A., Jacquier, A., Teichmann, J. (Eds.), Large Deviations and Asymptotic Methods in Finance. Springer-Verlag, Berlin, pp. 1–35.

Hobson, D., 2010. Comparison results for stochastic volatility models via coupling. Finance Stoch. 14(1), 129–152.

Howison, S.D, Steinberg, M., 2007. A matched asymptotic expansions approach to continuity corrections for discretely sampled options. Part 1: barrier options. Appl. Math. Finance 14(1), 63–89.

Ilhan, A., Jonsson, M., Sircar, R., 2004. Singular perturbations for boundary value problems arising from exotic options. SIAM J. Appl. Math. 64(4), 1268–1293.

Jiang, L., 2005. Mathematical Modelling and Methods of Option Pricing. World Scientific, Singapore.

Kevorkian, J., Cole, J.D., 1996. Multiple Scale and Singular Perturbation Methods. Springer, New York.

Knüsel, L., Bablok, B., 1996. Computation of the noncentral gamma distribution. SIAM J. Sci. Comput. 17(5), 1224-1231.

Kwok, Y.K., Wu, L., Yu, H., 1998. Pricing multi-asset options with an external barrier. Int. J. Theor. Appl. Finan. 1(4), 523–541.

Larguinho, M., Dias, J., Braumann, C., 2013. On the computation of option prices and Greeks under the CEV model. Quant. Finance 13(6), 907–917.

Lesniewski, A., 2009. Notes on the CEV model. Working paper.

Li, C., 2013. Maximum-likelihood estimation for diffusion processes via closed-form density expansions. Ann. Statist. 41(3), 1350–1380.

Li, C., 2014. Closed-form expansion, conditional expectation, and option valuation. Math. Oper. Res. 39(2), 487–516.

Linetsky, V., 2007. Spectral methods in derivatives pricing. Handbooks in Operations Research & Management Science, 15, 223–299.

Musiela, M., Rutkowski, M., 2004. Martingale Methods in Financial Modelling. Springer, New York.

Obłój, J., 2008. Fine-tune your smile: Correction to Hagan et al. Wilmott Magazine 35, 102–104.

Paulot, L., 2015. Asymptotic implied volatility at the second order with application to the SABR model, in: Friz, P., Gatheral, J., Gulisashvili, A., Jacquier, A., Teichmann, J. (Eds), Large Deviations and Asymptotic Methods in Finance. Springer-Verlag, Berlin, pp. 37-69.

Penev, S., Raykov, T., 2000. A Wiener germ approximation of the noncentral chi square distribution and of its quantiles. Comp. Stat. 15, 219–228.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C: The Art of Scientific Computing, secdon ed. Cambridge University Press, Cambridge.

Polyanin, D., 2001. Handbook of Linear Partial Differential Equations for Engineers and Scientists. Chaperman & Hall/CRC.

Rebonato, R., McKay, K., White, R., 2009. The SABR/LIBOR Market Model: Pricing, Calibration and Hedging for Complex Interest-Rate Derivatives. John Wiley & Sons.

Sankaran, M., 1963. Approximations to the non-central chi-square distribution. Biometrika 50, 199-204.

Schroder, M.S., 1989. Computing the constant elasticity of variance option pricing formula. J. Finance 44(1), 211–219.

Shreve, S.E., 2004. Stochastic Calculus for Finance II: Continuous-Time Models. Springer Science & Business Media.

Temme, N.E., 1993. Asymptotic and numerical aspects of the noncentral chi-square distribution. Computers Math.

Applic. 25(5), 55–63.

Watanabe, S., 1987. Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. Ann. Probab. 15, 1–39.

Widdicks, M., Duck, P., Andricopoulos, A., Newton, D., 2005. The Black-Scholes equation revisied: asymptotic expansions and singular perturbations. Math. Finance 15(2), 373–391.

Yang, N., Wan, X., 2016. First passage times of the SABR model: asymptotics and application. SSRN working paper #2847553.