

MASS AT ZERO IN THE UNCORRELATED SABR MODEL AND IMPLIED VOLATILITY ASYMPTOTICS

ARCHIL GULISASHVILI, BLANKA HORVATH, AND ANTOINE JACQUIER

ABSTRACT. We study the mass at the origin in the uncorrelated SABR stochastic volatility model, and derive several tractable expressions, in particular when time becomes small or large. As an application—in fact the original motivation for this paper—we derive small-strike expansions for the implied volatility when the maturity becomes short or large. These formulae, by definition arbitrage free, allow us to quantify the impact of the mass at zero on existing implied volatility approximations, and in particular how correct/erroneous these approximations become.

1. INTRODUCTION

The stochastic alpha, beta, rho (SABR) model introduced by Hagan, Kumar, Lesniewski and Woodward in [25, 27] is now a key ingredient—and has become an industry standard—on interest rates markets [2, 4, 7, 43]. It is defined by the pair of coupled stochastic differential equations

$$(1.1) \quad \begin{aligned} dX_t &= Y_t X_t^\beta dW_t, & X_0 &= x_0 > 0, \\ dY_t &= \nu Y_t dZ_t, & Y_0 &= y_0 > 0, \\ d\langle Z, W \rangle_t &= \rho dt, \end{aligned}$$

where $\nu > 0$, $\rho \in (-1, 1)$, $\beta \in (0, 1)$, and W and Z are two correlated Brownian motions on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. Its popularity arose from a tractable asymptotic expansion of the implied volatility (derived in [25]), and from its ability to capture the observed volatility smile; calibration therefore being made easier using the aforementioned expansion. In today's low interest rate and high volatility environment, the implied volatility obtained by this very expansion can however yield a negative density function for the price process X in (1.1), therefore exhibiting arbitrage.

This problem of negative density in low interest-rate environments has been directly addressed by Hagan et. al [26], Balland and Tran [7], and Andreasen and Høge [2], who proposed modifications of the original SABR model. There exist several refinements to the asymptotic formula itself: in [39] Oblój fine tunes the leading order, and Paulot [42] provides a second-order term. In certain parameter regimes the exact density has been derived for the absolutely continuous part (on $(0, \infty)$) of the distribution of X : in the uncorrelated case $\rho = 0$, formulae were obtained in [3, 29] by applying time-change techniques. The correlated case is much harder, and approximations have been derived using projection methods in [3, 4], using geometric tools in [24], and using PDE techniques [35, 41]. Barring computational costs, availability of the distribution of the SABR process is equivalent to computing any European prices. This however calls for a computation, not only on the continuous part of the distribution (on $(0, \infty)$), but also of its singular part at the origin. Absorbing boundary conditions at the origin ensure that the forward rate process X

Date: January 19, 2018.

2010 Mathematics Subject Classification. 60H30, 91G99, 91G60.

Key words and phrases. SABR model, asymptotic expansions, implied volatility.

The authors would like to thank Rama Cont and Josef Teichmann for initiating the series of ETH-Imperial College workshops, where this project initiated. BH would like to thank Leif Döring and Leonid Mytnik for stimulating discussions on time change techniques. BH acknowledges financial support from the SNF Early Postdoc Mobility Grant 165248. AJ acknowledges financial support from the EPSRC First Grant EP/M008436/1 and EPSRC Platform Grant EP/I019111/1. The numerical implementations have been carried out on the collaborative platform Zanadu (www.zanadu.io).

is a true martingale, and the singular part can hence accumulate mass, depending on the starting value of the process, the parameter configuration and the time horizon.

The original asymptotic formula typically loses accuracy for long-dated derivatives, when the CEV exponent β is close to zero, or when the volatility of volatility ν is large. The parameter β governs the dynamics of the smile, and small values thereof are usually chosen when the asymptotic formula fails, namely on markets where the forward rate is close to zero and for long-dated options [7, 25]. Indeed, it comes as no surprise that the original formula—which is an asymptotic expansion for small values of $\nu^2 T$ —breaks down for large maturities, but it is well known that the reasons for the inconsistencies of the SABR formula are subtler than that. What we highlight here is that the mass at zero can be held accountable for the irregularities in this case as well. While standard numerical methods proved reliable when the process remains strictly positive, computing the probability mass of the SABR model at the origin is a more delicate issue. Due to the singularity at the origin, usual regularity assumptions ensuring stability of numerical techniques (finite differences or Monte Carlo) are violated at this point, and a rigorous error analysis for these methods is not (yet) available. In addition, producing reference values becomes computationally intensive for short time scales.

Since much of the popularity of the SABR model is due to the tractability of its asymptotic formula, one should aim at preserving it while taking into account the mass at zero. The parameter sets $\rho = 0$ or $\beta = 0$ are the most tractable, and in fact (as observed in [13]) the only ones where certain advantageous regularity properties of the SABR process can be expected. We therefore concentrate here on the singular part of the distribution for these regimes, that is, we study the probability $\mathbb{P}(X_T = 0)$ and provide tractable formulae and asymptotic approximations. The relevance of these parameter configurations is emphasized by recent results [5], which suggest a so-called ‘mixture’ SABR (a combination of the $\rho = 0$ and $\beta = 0$ cases) approach to handle negative interest rates in an arbitrage-free way. From a modelling perspective, one may question the relevance of an absorbing boundary condition at zero in a financial context, where negative rates can actually occur. In fact, from a stochastic analysis perspective, when $\beta = 0$ there is no need to impose such a boundary condition. Remarkably however, as pointed out in [5], even in market conditions where interest rates become negative, the historical evolution of interest rates suggests that their dynamics follow processes whose probability distribution exhibit a singularity at the origin¹, which makes the computation of the mass at zero rates relevant for these market scenarios as well.

A further application is a direct approximation of the left wing of the implied volatility smile. In order to understand the small-strike behaviour of the SABR smile it is essential to determine the probability mass at the origin: asymptotic approximations of the implied volatility are available, not only for small and large maturities, but also for extreme strikes. Roger Lee’s celebrated Moment Formula [33]—subsequently refined by Benaim and Friz [9] and Gulisashvili [23]—relates the behaviour of the implied volatility $I_T(K)$ for small strike K and maturity T to the behaviour of the price process X around the origin. De Marco, Hillairet and Jacquier [12], and later Gulisashvili [21], showed that when the underlying distribution has an atom at zero, the small-strike behaviour of the implied volatility is solely determined by this mass, irrespective of the distribution of the process on $(0, \infty)$. We shall numerically confirm this in the (uncorrelated) SABR model, using approximations of the probability mass, in agreement with [6].

In Section 2 we derive explicit formulae for the mass at zero $\mathbb{P}(X_T = 0)$ in the SABR model for finite time as well as for large times in the uncorrelated case. Under this assumption, it is possible to decompose the distribution into a CEV component and an independent stochastic time change. Such time change techniques have been applied to the SABR model in the uncorrelated case in [3, 11, 29] to determine the exact distribution of the absolutely continuous part of the distribution on $(0, \infty)$. Therefore, our formulae complement these by providing the singular part of the distribution (see [28, 45] for more details about time change techniques in stochastic volatility models). In Section 2.2 and Section 2.3, we derive asymptotic expansions for the density of time-changed Brownian motion—inspired by the works of Borodin and Salminen [10], Gerhold [18] and

¹That is, rates ‘stick’ to zero for certain periods of time, see [5] for more details.

Matsumoto and Yor [37]—which we use to derive the behaviour of the atom at the origin for short and large times. Finally, in Section 3, we use these results to determine the left wing (small strikes) of the SABR implied volatility. Using the formulae provided in [12, 21], we highlight the fact that some of the widely used expansions exhibit arbitrage in the left wing, and propose a way to regularise them in this arbitrageable region.

2. MASS AT ZERO IN THE UNCORRELATED SABR MODEL

The price process X in (1.1) is a martingale [32, Remark 2]. If we consider X on the state space $[0, \infty)$, the origin, which can be attained, has to be absorbing [30, Chapter III, Lemma 3.6]. For two functions f and g , we shall write $f(z) \sim g(z)$ as z tends to zero whenever $\lim_{z \rightarrow 0} f(z)/g(z) = 1$.

2.1. The decomposition formula for the mass. In the case where the correlation coefficient ρ is null, the mass at the origin can be computed semi-explicitly. Conditioning on the path of the volatility process Y , the resulting process $\hat{X}$ satisfies the CEV stochastic differential equation

$$d\hat{X}_t = \hat{Y}_t \hat{X}_t^\beta dW_t,$$

starting from $\hat{X}_0 = x_0$, where $\hat{Y}$ is a deterministic time-dependent volatility coefficient, and represents, for fixed $\omega \in \Omega$, a realisation of the paths of Y . Consider now the simple CEV equation $d\tilde{X}_t = \tilde{X}_t^\beta dW_t$ starting from x_0 , and set

$$\hat{G}_t := \frac{\hat{X}_t^{2(1-\beta)}}{(1-\beta)^2} \quad \text{and} \quad \tilde{G}_t := \frac{\tilde{X}_t^{2(1-\beta)}}{(1-\beta)^2}.$$

Then $\hat{G}_t = Z_{\int_0^t \hat{Y}_s^2 ds}$, where Z is a Bessel process satisfying the SDE [29, Subsection 1.1]

$$dZ_t = \frac{1-2\beta}{1-\beta} dt + 2\sqrt{|Z_t|} dW_t, \quad Z_0 = \frac{x_0^{2(1-\beta)}}{(1-\beta)^2}.$$

By Itô's formula, the process $\tilde{G}$ solves the same SDE, so that $Z = \tilde{G}$, and therefore $\hat{X} = \tilde{X}_{\int_0^\cdot \hat{Y}_s^2 ds}$. It follows that X can be obtained from $\tilde{X}$ using the stochastic time change $t \mapsto \int_0^t Y_s^2 ds$, namely $X_t = \tilde{X}_{\int_0^t Y_s^2 ds}$. Since this time change is independent of $\tilde{X}$, one can decompose the mass at zero of the SABR model into that of the CEV component at zero and the density of the time change:

$$(2.1) \quad \mathbb{P}(X_t = 0) = \int_0^\infty \mathbb{P}(\tilde{X}_r = 0) \mathbb{P}\left(\int_0^t Y_s^2 ds \in dr\right) dr,$$

where the mass at zero in the CEV model is given by (see [12] or [31, Section 6.4.1])

$$(2.2) \quad \mathbb{P}(\tilde{X}_r = 0) = 1 - \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right),$$

with Γ , the normalised lower incomplete Gamma function: $\Gamma(v, z) \equiv \Gamma(v)^{-1} \int_0^z u^{v-1} e^{-u} du$.

Remark 2.1. If $\beta \in [1/2, 1)$ in (1.1), the origin is naturally absorbing, and the mass at zero is given by (2.2). When $\beta \in [0, 1/2)$, the solution to (1.1) is not unique, and a boundary condition at the origin has to be imposed. Should one consider the origin to be reflecting, the transition density would then become norm preserving, and no mass at the origin would be present. However, it is easy to see that there is an arbitrage opportunity if the origin is reflecting. Formula (2.2) carries over to the case $\beta \in [0, 1/2)$ when the origin is assumed to be absorbing, which we shall always consider from now on. This is of course in line with [30, Chapter III, Lemma 3.6], mentioned above, which states that the origin has to be absorbing for a non-negative supermartingale.

Since for each $s \geq 0$, Y_s is lognormally distributed, we can write

$$(2.3) \quad \mathbb{P}\left(\int_0^t Y_s^2 ds \in dr\right) = \mathbb{P}\left(\int_0^t \exp\left(2\nu Z_s^{(-\nu/2)}\right) ds \in d\tilde{r}\right),$$

where $\tilde{r} := \frac{r}{y_0^2}$, $Z_s^{(-\nu/2)} := Z_s - \frac{1}{2}\nu s$; the density of this functional is given by [10, Formula 1.10.4]

$$(2.4) \quad \mathbb{P}\left(\int_0^t e^{2\nu Z_s^{(-\nu/2)}} ds \in d\tilde{r}\right) = \frac{2^{1/4}\sqrt{\nu}}{\tilde{r}^{3/4}} \exp\left(-\frac{\nu^2 t}{8} - \frac{1}{4\nu^2 \tilde{r}}\right) m_{2\nu^2 t}\left(-\frac{3}{4}, \frac{1}{4\nu^2 \tilde{r}}\right) d\tilde{r},$$

where the function m is defined as [10, page 645]:

$$(2.5) \quad m_y(\mu, z) \equiv \frac{8z^{3/2}\Gamma(\mu + \frac{3}{2})e^{\frac{\pi^2}{4y}}}{\pi\sqrt{2\pi y}} \int_0^\infty e^{-z \cosh(2u) - \frac{1}{y}u^2} M\left(-\mu, \frac{3}{2}, 2z \sinh(u)^2\right) \sinh(2u) \sin\left(\frac{\pi u}{y}\right) du,$$

and where the Kummer function M reads

$$(2.6) \quad M(a, b, x) \equiv 1 + \sum_{k=1}^\infty \frac{a(a+1)\dots(a+k-1)x^k}{b(b+1)\dots(b+k-1)k!}.$$

2.2. Small-time asymptotics. We now study the behaviour of the mass at zero $\mathbb{P}(X_t = 0)$ as time becomes small. The main challenge is to provide a short-time asymptotic formula for the density of the time change process, for which standard expansion techniques are not applicable. The additive functional arising from the density of an integral of the exponential of Brownian motion often appears in the pricing of Asian options and is of interest on its own. This density is notoriously difficult to evaluate in small time, due to a highly oscillating factor connected to the Hartman-Watson distribution [36, 37] and [22, Section 4.6]. These numerical issues are discussed in [8], and Gerhold [18] used saddlepoint methods to provide short-time estimates. Because of the time change and the complexity of the Kummer function appearing in the integrand, small-time asymptotics of the mass at zero cannot be estimated directly. Instead, we use an inverse Laplace transform approach, inspired by [18], to provide small-time asymptotic estimates for the density of the time change. From (2.4) and (2.5), we introduce the notation $y := 2\nu^2 t$, and we shall alternate between the two notations without ambiguity in order to simplify some of the formulations below.

Remark 2.2. For $\varpi := 1/y$, the function m has the form $m_\varpi(\cdot) = c_\varpi \int_0^\infty e^{-u^2 \varpi} f_\varpi(u) du$, for some c_ϖ and f_ϖ . One might be tempted to use a standard Laplace method to determine the behaviour of m_ϖ as ϖ tends to infinity. However, at the saddlepoint $u^* = 0$, attained at the left boundary of the integration domain, all the derivatives of the function f_ϖ —appearing as coefficients of the expansion—are null, and the method does not apply.

We now formulate one of the main results of the paper, which characterises the small-time behaviour of the mass at zero in the uncorrelated SABR model. For every $r, y > 0$, let u_y denote the largest (positive) solution to the equation

$$(2.7) \quad 2\mu - 1 + 4uy + 2 \log(z/2)\sqrt{u} - \sqrt{u} \log(u) = 0,$$

with $z := \frac{y_0^2}{4\nu^2 r}$. Clearly, u_y depends on r , but we shall omit this dependence in the notation. Set

$$(2.8) \quad M_y := \frac{\log(u_y)}{16u_y^{3/2}} - \frac{\alpha}{8u_y^{3/2}} + \frac{1 - 2\mu}{8u_y^2}.$$

The following theorem provides a short-time asymptotic estimate for the density of the integrated variance process given in (2.3) and (2.4).

Theorem 2.3. *As y (equivalently t) tends to zero, we have (recall that $y = 2\nu^2 t$)*

$$\mathbb{P}\left(\int_0^t Y_s^2 ds \in dr\right) \sim \frac{y_0^{3/2} e^{5/4}}{r^{5/4} 2^{7/4} \sqrt{\nu\pi}} \exp\left(-\frac{y}{16} - \frac{y_0^2}{4\nu^2 r}\right) \exp\left[\frac{\log(u_y)}{2} \left(\mu - \frac{1}{2}\right) - u_y y + \sqrt{u_y}\right] \frac{dr}{\sqrt{M_y}}.$$

The proof of the theorem relies on the representation (2.4) of the integrated variance, which has the crucial, strongly oscillating factor m_y in (2.5). Therefore, the key step in proving Theorem 2.3 is the derivation of a small-time asymptotic expansion for m_y , which we deliver in the following proposition. The technical part of the proof of this proposition is delegated to the Appendix A.1 to lighten the flow of the paper.

Proposition 2.4. *As y (equivalently t) tends to zero, the function m_y in (2.5) satisfies*

$$m_y(\mu, z) \sim \frac{\sqrt{z} \exp\left(\frac{1}{2} - \mu\right)}{2\pi} \exp\left[\frac{\log(u_y)}{2} \left(\mu - \frac{1}{2}\right) - u_y y + \sqrt{u_y}\right] \sqrt{\frac{\pi}{M_y}}.$$

Remark 2.5. The proof of the proposition uses saddlepoint analysis. The saddlepoint u_y is the solution to (2.7), but does not admit a closed-form expression; however, as seen in the proof, it is possible to expand it as y tends to zero to obtain

$$m_y(\mu, z) = \frac{\sqrt{z} |\log(y)|}{2\sqrt{\pi}} \exp\left\{-\frac{\log(y)^2}{4y} + \frac{|\log(y)|}{2y} + \left(\frac{1}{2} - \mu\right) \left[1 - \log\left(\frac{|\log(y)|}{2y}\right)\right]\right\} \left[\frac{1}{y^{3/2}} + \mathcal{O}\left(y^{3/2}\right)\right].$$

It appears tempting and natural to use the above closed-form expansion directly for the approximation of (2.5) in (2.4), but numerical computations show that this expression is not very precise and should therefore not be used. On the other hand, the expansion derived in Proposition 2.4, gives an accurate approximation of (2.5) for the computation of (2.4). This underlines the relevance of Proposition 2.4 for practical purposes.

With the approximation formula for the integrated variance in Theorem 2.3 at hand, we are now in a position to formulate a numerical scheme for the mass at zero in the uncorrelated SABR model (1.1), based on the decomposition (2.1). Due to the subtleties caused by the fluctuations of the function m_y for small y discussed in Remark 2.5, the approximation of the mass at zero $\mathbb{P}(X_t = 0)$ as t tends to zero is a delicate task. We construct here a hybrid numerical scheme, based on Theorem 2.3, that delivers a better approximation (in terms of speed and accuracy) than any of its components on their own². In the following mixing scheme, the expansion of Theorem 2.3 is combined with a lognormal approximation of the integrated variance I_t from (2.3) at small values of $r \in (0, \infty)$. In the lognormal scheme we approximate $I_t =: y_0^2 J_t \approx y_0^2 \Theta_t$, where Θ_t is a log-normal random variable whose first and second moments match those of J_t , for $t > 0$. In the following numerical scheme the composition of the approximation methods is parametrised by a mixing parameter $\alpha_t > 0$, depending on the time horizon $t > 0$ and on the parameters β, ν and y_0 of the (uncorrelated) model, and is optimised with respect to approximation accuracy:

$$(2.9) \quad \mathbb{P}(X_t = 0) \approx \int_0^\infty \mathbb{P}(\tilde{X}_r = 0) \left((1 - \alpha_t) f_1(r; y_0, \nu) + \alpha_t f_2(r; y_0, \nu) \right) dr,$$

where $f_1(r; y_0, \nu)$ denotes the approximation of the density of I_t (2.3) via Theorem 2.3:

$$(2.10) \quad \begin{aligned} f_1(r; y_0, \nu) &\equiv \exp\left\{\frac{\log(u_y)}{2} \left(\mu - \frac{1}{2}\right) - u_y y + \sqrt{u_y}\right\} \frac{g(r)}{\sqrt{M_y}}, \\ \text{with } g(r) &\equiv \frac{y_0^{3/2} e^{5/4}}{2^{7/4} \sqrt{\nu \pi}} \exp\left(-\frac{\nu^2 t}{8}\right) \frac{1}{r^{5/4}} \exp\left(-\frac{y_0^2}{4\nu^2 r}\right), \quad \text{for } r \in [0, \infty), \end{aligned}$$

and where $f_2(r; y_0, \nu)$ is the lognormal approximation of I_t (2.3) given by:

$$(2.11) \quad \begin{aligned} f_2(r; y_0, \nu) &\equiv \frac{1}{y_0^2} f_\Theta\left(\frac{r}{y_0^2}\right), \quad r \in [0, \infty), \\ \text{with } f_\Theta(z) &\equiv \frac{1}{z\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log(z) - \mu)^2}{2\sigma^2}\right\}, \quad z \in (0, \infty), \end{aligned}$$

where μ and σ^2 are the parameters of the lognormal proxy Θ_t . Since $\mathbb{E}(\Theta_t) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$ and $\mathbb{E}(\Theta_t^2) = \exp(2\mu + 2\sigma^2)$, and $\mathbb{E}(J_t) = \frac{1}{\nu^2} (e^{\nu^2 t} - 1)$ and $\mathbb{E}(J_t^2) = \frac{1}{15\nu^4} (e^{6\nu^2 t} - 6e^{\nu^2 t} + 5)$, then matching the first two moments of Θ_t and J_t yields

$$\mu_t = \log\left(\frac{\mathbb{E}(J_t)^2}{\sqrt{\mathbb{E}(J_t^2)}}\right) \quad \text{and} \quad \sigma_t^2 = \log\left(\frac{\mathbb{E}(J_t^2)}{\mathbb{E}(J_t)^2}\right).$$

An additional reason is that for given t as r becomes small, the saddlepoint equation (2.7) can have

²We thank the anonymous referee for making us aware of this improvement.

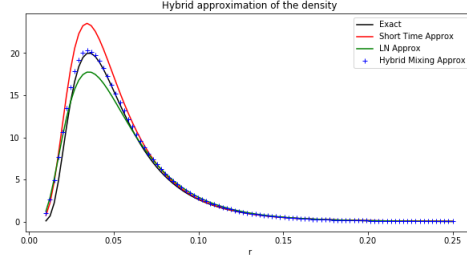


FIGURE 1. The advantage of mixing these two approximation methods becomes clear from the graphs in the figure above. While the lognormal approximation (green line) underestimates the exact density (line labelled *Exact*) of the integrated variance I_t (2.3), the approximation via Theorem 2.3 (red line) slightly overestimates it. The graphs display approximations of the density for the parameters $(y_0, \nu) = (0.6, 1)$ and $\beta = 0.3$ at maturity $T = 0.3$. We used the value $\alpha = 0.7$ for this approximation.

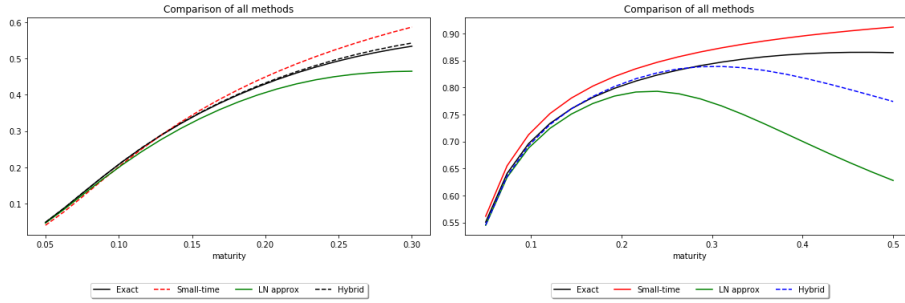


FIGURE 2. The figure illustrates the accuracy of the numerical scheme (2.9) for the short-time approximation of the mass at zero and a comparison with other methods. In the images above, the scheme (2.9) is illustrated as the line labelled as *Hybrid* scheme, which performs remarkably well on time horizons up to $T = 0.3$. The images show approximations of the mass with parameters $(y_0, \nu) = (0.6, 1)$ and (left) $\beta = 0.5$ up to $T = 0.3$ such as (right) $\beta = 0.3$ up to $T = 0.5$.

bifurcations in the number of solutions, which can lead to instabilities in integration. Switching to the lognormal approximation of the integrated variance at very small values of $r > 0$ bypasses this problem at the cost of a small approximation error. In a second step we also determine the optimal switching point in the sense of accuracy where we test against a direct quadrature of (2.1).

2.3. Large-time asymptotics. We now concentrate on the large-time behaviour of the mass at zero in the uncorrelated SABR model. From a practical point of view, this case is less relevant as data suggests the presence of a leverage effect, usually captured by negative values of the correlation parameter. However this leverage, to a certain extent, can also be partially generated with non-trivial values of the parameter β . From [10, Formula 1.8.4, page 612], the formula

$$\mathbb{P} \left(\int_0^\infty \exp \left(2\nu Z_s^{(-\nu/2)} \right) ds \in d\tilde{r} \right) = \frac{\tilde{r}^{-3/2}}{\nu\sqrt{2\pi}} \exp \left(-\frac{1}{2\nu^2\tilde{r}} \right) d\tilde{r}$$

holds, so that the decomposition (2.1) together with (2.3) imply that (recall that $\tilde{r} = \frac{r}{y_0^2}$)

$$\begin{aligned} \mathbb{P}_\infty &:= \lim_{t \uparrow \infty} \mathbb{P}(X_t = 0) = \frac{y_0}{\nu\sqrt{2\pi}} \int_0^\infty \left[1 - \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right) \right] r^{-3/2} \exp\left(-\frac{y_0^2}{2\nu^2 r}\right) dr \\ (2.12) \quad &= 1 - \frac{y_0}{\nu\sqrt{2\pi}} \int_0^\infty \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right) r^{-3/2} \exp\left(-\frac{y_0^2}{2\nu^2 r}\right) dr. \end{aligned}$$

When $\beta = 0 (= \rho)$, the SABR model (1.1) reduces to a Brownian motion on the hyperbolic plane (up to a deterministic time change), and a simple computation shows that (2.12) simplifies to

$$\mathbb{P}_\infty|_{\beta=0} = 1 - \frac{2}{\pi} \arctan\left(\frac{\nu x_0}{y_0}\right).$$

When $\beta \neq 0$, the integral in (2.12) does not have a closed-form expression. Expanding the exponential factor for small y_0 , we can however write, for any $n \in \mathbb{N}$, the n th-order approximation

$$\begin{aligned} \mathbb{P}_\infty^{(n)} &:= \int_0^\infty \left[1 - \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right) \right] \frac{y_0}{\nu r^{3/2} \sqrt{2\pi}} \sum_{k=0}^n \frac{1}{k!} \left(-\frac{y_0^2}{2\nu^2 r}\right)^k dr \\ &= \sum_{k=0}^n \frac{y_0^{2k+1}}{k! \nu \sqrt{2\pi}} \left(-\frac{1}{2\nu^2}\right)^k \int_0^\infty \left[1 - \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right) \right] r^{-(k+3/2)} dr \\ &= \frac{2y_0(1-\beta)}{\Gamma\left(\frac{1}{2(1-\beta)}\right) \nu \sqrt{\pi} x_0^{1-\beta}} \sum_{k=0}^n \frac{(-1)^k}{k!} \left(\frac{y_0^2(\beta-1)^2}{\nu^2 x_0^{2(1-\beta)}}\right)^k \frac{\Gamma\left(k+1+\frac{\beta}{2-2\beta}\right)}{(1+2k)}. \end{aligned}$$

Note in particular that

$$(2.13) \quad \mathbb{P}_\infty^{(0)} = \frac{2\Gamma\left(1+\frac{\beta}{2-2\beta}\right) y_0(1-\beta)}{\Gamma\left(\frac{1}{2-2\beta}\right) \nu \sqrt{\pi} x_0^{1-\beta}}.$$

When r tends to infinity, the integrand clearly converges to zero fast enough. Using the properties of Gamma functions in [1, Chapter 6], the asymptotic behaviour

$$1 - \Gamma\left(a, \frac{1}{r}\right) \sim \frac{r^{1-a} \exp(-1/r)}{\Gamma(a)}$$

holds as r tends to zero, ensuring that the integral is well defined for all $n \in \mathbb{N}$. Theorem 2.6 below shows how well (and when) the sequence $\mathbb{P}_\infty^{(n)}$ approximates the mass at zero $\mathbb{P}_\infty$. Using the Taylor formula with Lagrange's form of the remainder, we obtain

$$\exp\left(-\frac{y_0^2}{2\nu^2 r}\right) = \sum_{k=0}^n (-1)^k \frac{1}{k!} \left(\frac{y_0^2}{2\nu^2 r}\right)^k + \frac{(-1)^{n+1}}{(n+1)!} e^{-\theta} \left(\frac{y_0^2}{2\nu^2 r}\right)^{n+1},$$

for some $\theta \in (0, y_0^2/(2\nu^2 r))$. Therefore,

$$(2.14) \quad \left| \exp\left\{-\frac{y_0^2}{2\nu^2 r}\right\} - \sum_{k=0}^n (-1)^k \frac{1}{k!} \left(\frac{y_0^2}{2\nu^2 r}\right)^k \right| \leq \frac{1}{(n+1)!} \left(\frac{y_0^2}{2\nu^2 r}\right)^{n+1}.$$

For any $n \geq 0$, set

$$(2.15) \quad b_n := \frac{2y_0(1-\beta)}{\Gamma\left(\frac{1}{2(1-\beta)}\right) \nu \sqrt{\pi} x_0^{1-\beta}} \left(\frac{y_0^2(\beta-1)^2}{\nu^2 x_0^{2(1-\beta)}}\right)^n \frac{\Gamma\left(n+1+\frac{\beta}{2-2\beta}\right)}{n!(1+2n)},$$

so that from (2.14), (2.15), and the definitions of $\mathbb{P}_\infty$ and $\mathbb{P}_\infty^{(n)}$, it follows that, for any $n \geq 0$,

$$(2.16) \quad \mathbb{P}_\infty^{(n)} = \sum_{k=0}^n (-1)^k b_k \quad \text{and} \quad \left| \mathbb{P}_\infty - \mathbb{P}_\infty^{(n)} \right| \leq b_{n+1}.$$

Theorem 2.6. *The following statements hold for the sequence $\mathbb{P}^{(n)}$ in (2.16):*

- (i) *if $y_0^2(\beta - 1)^2 > \nu^2 x_0^{2(1-\beta)}$, or $y_0^2(\beta - 1)^2 = \nu^2 x_0^{2(1-\beta)}$ and $\frac{2}{3} \leq \beta < 1$, then the sequence $(\mathbb{P}_\infty^{(n)})_{n \geq 0}$ diverges, and hence cannot be an approximation to the mass at zero $\mathbb{P}_\infty$;*
- (ii) *if $y_0^2(\beta - 1)^2 < \nu^2 x_0^{2(1-\beta)}$, or $y_0^2(\beta - 1)^2 = \nu^2 x_0^{2(1-\beta)}$ and $0 \leq \beta < \frac{2}{3}$ then*

$$(2.17) \quad \mathbb{P}_\infty = \mathbb{P}_\infty^{(n)} + \mathcal{O} \left(n^{-1 + \frac{\beta}{2-2\beta}} \exp \left(-n \log \left(\frac{\nu^2 x_0^{2(1-\beta)}}{y_0^2(\beta - 1)^2} \right) \right) \right), \quad \text{as } n \text{ tends to infinity.}$$

Remark 2.7. Remember that x_0 denotes the initial value of the stock price or interest rate. For all practical and sensible values of the parameters, condition (ii) in the theorem is always in force.

Proof. From (2.15), Stirling's formula for the Gamma function yields, as k tends to infinity,

$$(2.18) \quad b_k \sim \frac{y_0(1-\beta)}{\Gamma\left(\frac{1}{2(1-\beta)}\right) \nu \sqrt{\pi} x_0^{1-\beta}} k^{-1 + \frac{\beta}{2-2\beta}} \left(\frac{y_0^2(\beta - 1)^2}{\nu^2 x_0^{2(1-\beta)}} \right)^k,$$

From (2.16), if the conditions of Theorem 2.6(i) hold, the general term of the series $\sum_{k=0}^{\infty} (-1)^k b_k$ does not tend to zero, and the sequence $\mathbb{P}_\infty^{(\cdot)}$ diverges. If the conditions of Theorem 2.6(ii) hold, then (2.16) and (2.18) imply (2.17), which completes the proof of Theorem 2.6. $\square$

For practical purposes, depending on the conditions for convergence of the sequence $(\mathbb{P}_\infty^{(n)})_{n \geq 0}$ in Theorem 2.6, it may or may not be useful to use directly the integral form (2.12). For $(y_0, \nu, \beta, x_0) = (0.1, 1.0, 0.2, 0.2)$ (for which convergence holds), the mass at zero in this case is $\mathbb{P}_\infty = 20.833\%$. Using Theorem 2.6, we obtain the following tables the table below computes the error using the sequence $(\mathbb{P}_\infty^{(n)})$:

	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
$ \mathbb{P} - \mathbb{P}_\infty^{(n)} $	6.43E-3	3.41E-4	2.13E-05	1.43E-06	1.01E-07	7.29E-09
Computation time (in seconds)	6.8E-05	8.6E-05	1.3E-4	1.9E-4	2.2E-4	2.6E-4

The first table computes the error obtained by using the sequence $(\mathbb{P}_\infty^{(n)})$, while the second one provides the values of the integral (2.12), which were obtained by using the Python scipy toolpack for quadrature; the integral is truncated at some arbitrary value $R > 0$.

	$R = 20$	$R = 40$	$R = 60$	$R = 80$	$R = 100$	$R = 120$
Absolute error	2.33E-4	1.07E-4	6.77E-05	4.90E-05	3.81E-05	3.11E-05
Computation time (in seconds)	7.6E-3	7.9E-3	8.9E-3	9.2E-3	9.6E-3	9.9E-3

These results suggest that convergence of the series expansion is extremely fast. In particular, even the limit (2.13), with $n = 0$, yields a very accurate result, which allows for a simple interpretation of the impact of each parameter of the model on the large-time mass at the origin.

2.3.1. Large-time numerics. We provide some numerics for the large-time mass at zero derived in (2.12). In particular, we observe the influence of the parameter β (Figure 3) as well as that of the starting point x_0 (Figure 3) of the uncorrelated SDE (1.1). As β tends to one (from below), the mass at zero is diminishing, even for arbitrarily small values of x_0 . Likewise, as the initial value x_0 increases, the mass at the origin decreases even for $\beta = 0$. We shall further comment on the importance of the mass at the origin in financial modelling in Section 3 below.

We include for comparison a sample of values for the mass at zero obtained by Monte Carlo simulations with $M = 1000$ and $M = 2000$ paths, and the corresponding computation times, for different time horizons T . The results suggest that the 'large-time' regime is already achieved for maturities equal to 15 years. An explanation for this phenomenon is provided in [13, Section 4]. As in Section 2.3 above, we used the parameters $(y_0, \nu, \beta, x_0) = (0.1, 1.0, 0.2, 0.2)$, for which the exact mass at zero is $\mathbb{P}_\infty = 20.833\%$.

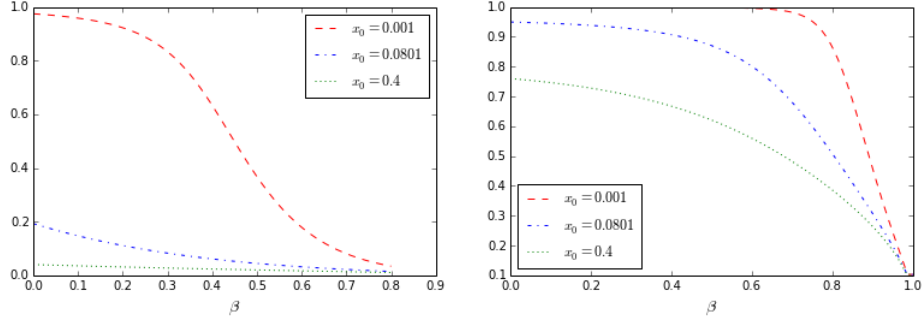


FIGURE 3. Influence of β on the large-time mass at zero in the uncorrelated SABR model with $(y_0, \nu) = (0.015, 0.6)$ (left) and $(y_0, \nu) = (0.1, 1)$ (right).

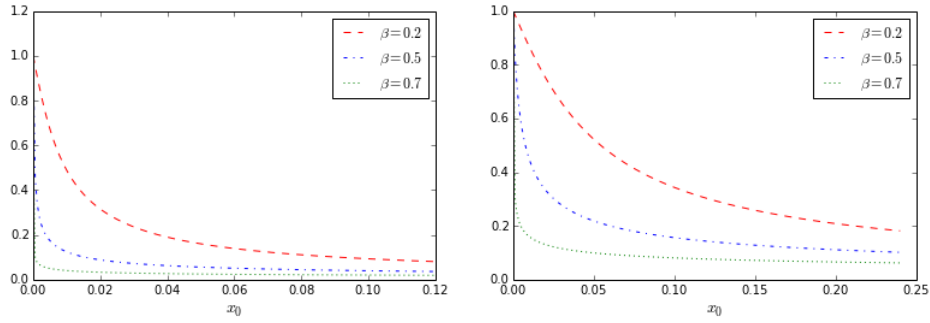


FIGURE 4. Influence of the initial value x_0 on the large-time mass at zero with $(y_0, \nu) = (0.015, 0.6)$ (left) and $(y_0, \nu) = (0.1, 1)$ (right). This gives a numerical interpretation of ‘feeling the boundary’: as we start the diffusion far enough from the origin, the mass at zero becomes small.

	$T = 10$	$T = 15$	$T = 20$	$T = 30$	$T = 50$	$T = 100$
Monte Carlo mass (M=1000)	0.1889	0.2020	0.2100	0.2110	0.2050	0.2090
Computation time (in seconds)	3.222	3.185	3.221	3.179	3.163	3.178

	$T = 10$	$T = 15$	$T = 20$	$T = 30$	$T = 50$	$T = 100$
Monte Carlo mass (M=2000)	0.2100	0.2075	0.2050	0.2100	0.2065	0.2175
Computation time (in seconds)	6.4437	6.8386	6.4805	7.6308	6.6587	6.372

When comparing our asymptotics with Monte Carlo results, one should keep in mind that, the numbers in the above table were obtained by an Euler scheme, and it is known [11] that in the critical region around zero, Monte Carlo methods are prone to a simulation bias. While Monte Carlo techniques have the unquestionable advantage that they can handle the correlated case and are widely used in quant libraries, precise convergence rates are not (yet) available. Recent advances have been made in [20, 34] proposing sophisticated and efficient Monte-Carlo simulation of the SABR model. In the development of new Monte Carlo techniques for the critical regime around zero, our results can serve as useful benchmarks, independent from the random seed.

3. IMPLIED VOLATILITY AND SMALL-STRIKE EXPANSIONS

The implied volatility is the Black-Scholes volatility parameter that allows to match observed (or computed) European option prices; it obviously depends on strikes and maturities (see for

example [16] for more details). Given a model, the classical route to compute the implied volatility is (i) to compute the price of the Call (or Put) option, and (ii) to invert the Black-Scholes formula. Both steps are numerically demanding, and scarcely provide insights on the behaviour of the implied volatility smile. Another route, which has motivated the use of asymptotic methods in finance, is to obtain closed-form expansions for the smile (for small/large maturities, or strikes); however, being asymptotic results, they may lose accuracy when some parameters are not small/large enough. The ‘classical’ approximation by Hagan et al [27] is such a formula, which has been used extensively by practitioners, despite exhibiting flaws—namely arbitrage—in some regions. This anomaly can in principle be fixed if one accounts for the accumulation of mass at zero due to the Dirichlet boundary condition. Let us recall a few (model-independent) results regarding small-strike asymptotics of the implied volatility. For any strike $K > 0$ and maturity $T > 0$, let us denote by $I_T(K)$ the implied volatility. In the presence of strictly positive mass at zero, the small-strike tail of the implied volatility satisfies [33]:

$$(3.1) \quad \limsup_{K \downarrow 0} \frac{I_T(K)}{\sqrt{|\log K|}} = \sqrt{\frac{2}{T}}.$$

This behaviour was recently refined by De Marco, Hillairet and Jacquier [12], and later by Gulisashvili [21]. Assuming that $\mathbb{P}(X_T \leq K) - \mathbb{P}(X_T = 0) = \mathcal{O}(|\log K|^{-3/2})$ as K tends to zero, De Marco, Hillairet and Jacquier [12, Proposition 3.1] derive the small-strike asymptotic formula

$$(3.2) \quad I_T(K) = \sqrt{\frac{2|\log K|}{T}} + \frac{\mathcal{N}^{-1}(m_T)}{\sqrt{T}} + \frac{(\mathcal{N}^{-1}(m_T))^2 + 2}{2\sqrt{2T|\log K|}} + \frac{(\mathcal{N}^{-1}(m_T))}{4|\log K|\sqrt{T}} + \mathcal{O}(|\log K|^{-3/2}),$$

where $m_T := \mathbb{P}(X_T = 0)$ is the mass at the origin, and $\mathcal{N}$ the Gaussian cumulative distribution function (an alternative formulation of (3.2) can be found in [21]).

3.1. Numerical illustration of the arbitrage-free tail expansion. In this subsection we apply results of [12] to test our small-strike asymptotic implied volatility formula (3.2) for the presence of arbitrage. Following [12], if a mass at the origin is present, the map $k \mapsto I_T(e^k)\sqrt{T/|k|}$, has to be bounded by $\sqrt{2}$ in order to avoid arbitrage. In Figure 3.1, we plot $k \mapsto I_T(e^k)\sqrt{T/|k|}$, the first and second order of (3.2) using (2.12) to compute the mass at zero. We consider two parameter sets, one for which the mass at zero is small, and another one yielding a large mass at the origin. The numerics visualise that our formulas (3.2) are indeed arbitrage free, with values well below the $\sqrt{2}$ threshold. We furthermore include a comparison to the original SABR formula (here, we implement it including the correction suggested by Oblój [39]). The numerics indicate a relationship between size of the mass at the origin and the level of violation of the no-arbitrage condition. The graphs visually quantify how ‘wrong’ Hagan’s expansion (using Oblój’s version) is for small strikes in the presence of a mass at the origin for the parameter set yielding a large mass at zero, and for the one that yields a small mass. As the mass becomes small, Hagan’s approximation becomes more accurate. This holds in particular as β approaches one, as indicated in Section 2.3.1 above. In the limit as $\beta = 1$, the mass becomes null.

3.2. Comparison with Antonov-Konikov-Spector [6]. In the uncorrelated case $\rho = 0$, Antonov, Konikov and Spector [6] derived the double integral formula for the price of a Call option:

$$\mathbb{E}(X_T - K)_+ = (X_0 - K)_+ + \frac{2\sqrt{X_0 K}}{\pi} \left\{ \int_{s_-}^{s_+} \frac{\sin(\eta\varphi(s))}{\sinh(s)} G(\nu^2 T, s) ds + \sin(\eta\pi) \int_{s_+}^{\infty} \frac{\exp(-\eta\psi(s))}{\sinh(s)} G(\nu^2 T, s) ds \right\},$$

where $\eta := 1/2(\beta - 1)$, $q := \frac{K^{1-\beta}}{1-\beta}$, $q_0 := \frac{X_0^{1-\beta}}{1-\beta}$, $s_{\pm} := \operatorname{arcsinh}\left(\frac{\nu}{y_0}|q \pm q_0|\right)$,

$$\varphi(s) := 2 \arctan \sqrt{\frac{\sinh(s)^2 - \sinh(s_-)^2}{\sinh(s_+)^2 - \sinh(s)^2}} \quad \text{and} \quad \psi(s) := 2 \operatorname{arctanh} \sqrt{\frac{\sinh(s)^2 - \sinh(s_+)^2}{\sinh(s)^2 - \sinh(s_-)^2}}.$$

The function G is defined as the integral

$$G(t, s) := \frac{2 \exp(-t/8)}{t^{3/2} \sqrt{\pi}} \int_s^{\infty} u \sqrt{\cosh(u) - \cosh(s)} \exp\left(-\frac{u^2}{2t}\right) du.$$

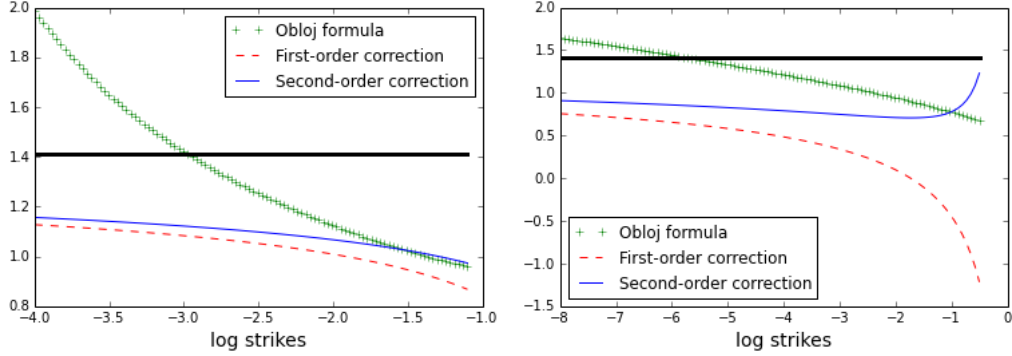


FIGURE 5. The black line marks the level $\sqrt{2}$. The parameters are $(\nu, \beta, \rho, x_0, y_0, T) = (0.3, 0, 0, 0.35, 0.05, 10)$ for the left plot, and $(\nu, \beta, \rho, x_0, y_0, T) = (0.6, 0.6, 0, 0.08, 0.015, 10)$ for the right graph. Obłój's implied volatility expansion clearly violates this upper bound in both cases. The large-time mass is equal to 28.3% for the left plot and 3.1% for the right one.

In Figure 6, we compare the smile obtained from the Antonov-Konikov-Spector's formula (computing the double integral and numerically inverting the Black-Scholes formula) and the closed-form tail formula (3.2) using the large-maturity mass at zero computed from (2.12). Following [3], we consider the following set of parameters: $(\nu, \beta, \rho, x_0, y_0, T) = (0.8, 0.1, 0, 0.1, 0.15, 20)$, for which the large-maturity mass at zero (2.12) is equal to 63%. In the other graphs we vary the value of the parameters β and y_0 between $\beta = 0.0$ and $\beta = 0.9$ such as y_0 between $y_0 = 0.15$ and $y_0 = 0.7$ to achieve different value of the probability mass. We also tested against different time horizons of the exact formulas, from $T = 2$ to $T = 20$ and the agreement remains robust on these horizons. We find good agreement between these formulas in all of these cases throughout different values of the mass as the following graphs illustrate.

4. CONCLUSION

The SABR model is a pillar of mathematical modelling on fixed income desks, but suffers from some issues in low interest rate environments, where the process can hit the origin with non-zero probability, creating, in most existing approximations (used by practitioners), arbitrage opportunities. In this paper, we endeavour to provide accurate estimates for this mass at zero in order (i) to quantify the error made by existing approximations, and (ii) to suggest an alternative parameterisation of the implied volatility smile for low strikes, ensuring arbitrage opportunities do not occur.

APPENDIX A. PROOFS OF SECTION 2

A.1. Proof of Proposition 2.4. Our proof is inspired by [18], which is based on an inverse Laplace transform approach. From [10, Page 645], the Laplace transform of the function m_y has a closed-form representation, namely, whenever $\mu > -3/2$ and $z > 0$,

$$m_y(\mu, z) = \mathcal{L}_u^{-1} \left(\frac{\Gamma(\mu + \frac{1}{2} + \sqrt{u})}{\Gamma(1 + 2\sqrt{u})} \mathfrak{M}_{-\mu, \sqrt{u}}(2z) \right),$$

where the function $\mathfrak{M}$ is related to the Kummer function M function via the identity

$$\mathfrak{M}_{n,m}(x) \equiv x^{m+1/2} \exp\left(-\frac{x}{2}\right) M\left(m - n + \frac{1}{2}, 2m + 1, x\right).$$

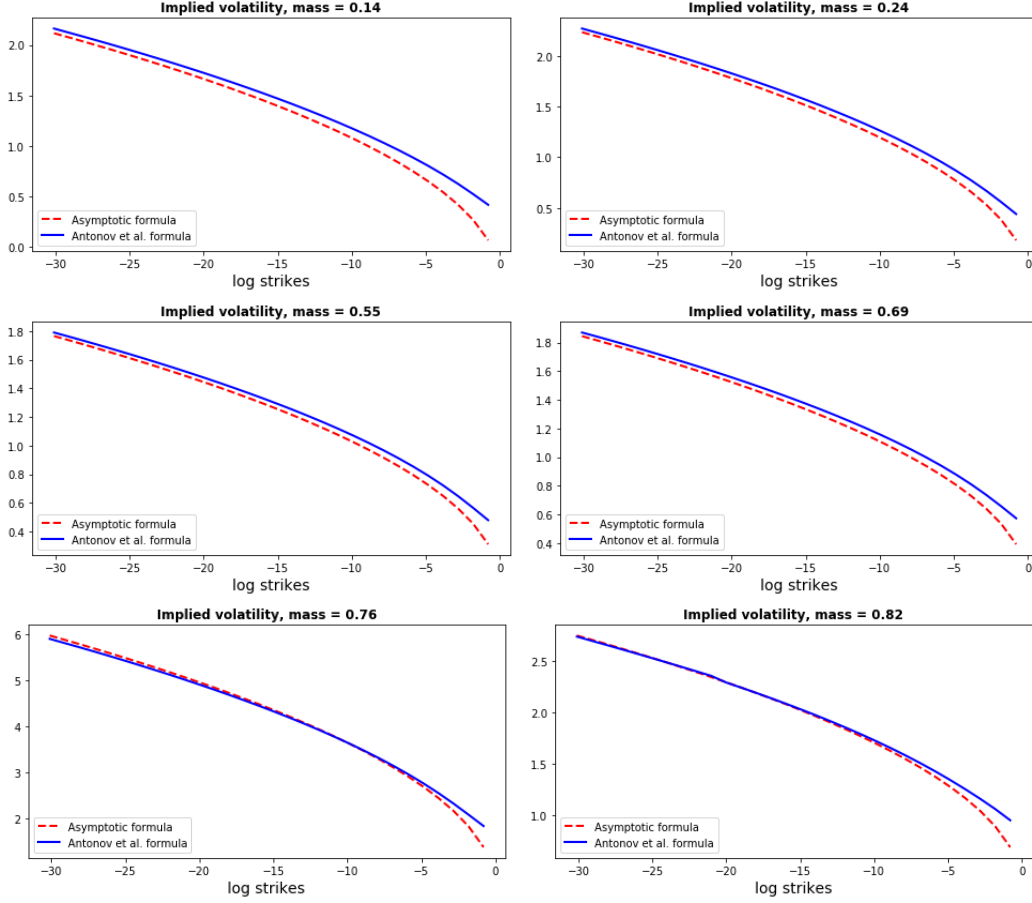


FIGURE 6. The graph suggests that there is good agreement between the two formulas despite completely orthogonal approaches: While the Antonov-Konikov-Spector formulas are computed from the knowledge of the exact *densities* of the distribution on the interval $X \in (0, \infty)$, our asymptotic formulas are computed solely from the probability *mass* at $X = 0$. Despite describing complementary parts of the probability distribution of the underlying price process X in (1.1), the two formulas show perfect agreement in the critical region of small strikes and good agreement throughout the entire range of strikes.

Therefore, we can write, for some $R \in \mathbb{R}$,

$$(A.1) \quad m_y(\mu, z) = \frac{e^{-z}}{2i\pi} \int_{R-i\infty}^{R+i\infty} e^{uy} \frac{\Gamma(\mu + \frac{1}{2} + \sqrt{u})}{\Gamma(1 + 2\sqrt{u})} (2z)^{\frac{1}{2} + \sqrt{u}} M\left(\mu + \frac{1}{2} + \sqrt{u}, 1 + 2\sqrt{u}, 2z\right) du.$$

When analysing the conditional density of the integrated exponential Brownian motion, Gerhold [18], as already mentioned, dealt with a similar integral. However, before being able to adapt his saddlepoint methodology to our case, we first need to deal with the Kummer function M . Since we wish to determine the behaviour of m_y as y (equivalently, t) tends to zero, we need to understand the limit of the integrand as u tends to infinity. The following asymptotic relations hold uniformly in v , as $v = \sqrt{u}$ tends to infinity:

$$(A.2) \quad \Gamma(1 + v) = \sqrt{2\pi} e^{-v} v^{v+1/2} [1 + \mathcal{O}(v^{-1})] \quad \text{and} \quad M\left(\mu + \frac{1}{2} + v, 1 + 2v, 2z\right) \sim e^z.$$

The first one is standard [38, Section 3.5]. As for the second one, the representation (2.6) yields

$$M\left(\frac{1}{2} + v + \mu, 1 + 2v, 2z\right) = \sum_{k=0}^{\infty} \gamma_k \frac{(2z)^k}{k!},$$

where, for $k \geq 0$,

$$\gamma_k := \frac{(\mu + v + \frac{1}{2}) \cdots (\mu + v + k - \frac{1}{2})}{(1 + 2v)(2 + 2v) \cdots (k + 2v)}.$$

Clearly $|\gamma_k| \leq 2^{-k}$ and $\gamma_k \sim 2^{-k}$ as v tends to infinity, and from [1, Formula 13.6.3], we have

$$M\left(\frac{1}{2} + v + \mu, 1 + 2v, 2z\right) \leq M\left(\frac{1}{2} + v, 1 + 2v, 2z\right) = \Gamma(1 + v) e^z \left(\frac{1}{2} z\right)^{-v} I_v(z),$$

where again I_v denotes the modified Bessel function of the first kind [10, Page 638], so that, using (A.2) and [18, Equation 9], we have, uniformly in v ,

$$\left| M\left(\frac{1}{2} + v + \mu, 1 + 2v, 2z\right) \right| \leq \Gamma(1 + v) e^z \left(\frac{z}{2}\right)^{-v} I_v(z) = e^z (1 + \mathcal{O}(v^{-1})).$$

Therefore the integrand in (A.1) reads, as u tends to infinity,

$$\begin{aligned} \Phi(u, y, z) &\equiv e^{uy} \frac{\Gamma(\mu + \frac{1}{2} + \sqrt{u})}{\Gamma(1 + 2\sqrt{u})} (2z)^{\frac{1}{2} + \sqrt{u}} M\left(\mu + \frac{1}{2} + \sqrt{u}, 1 + 2\sqrt{u}, 2z\right) \\ &\sim e^{v^2 y + v + z} z^{v + \frac{1}{2}} v^{\mu - v - \frac{1}{2}} 2^{-v} \\ &= \exp\left[v^2 y + \alpha v + \left(\mu - \frac{1}{2} - v\right) \log(v) + z + \frac{1}{2} \log(z)\right] =: \exp\left(\psi_y(u) + z + \frac{1}{2} \log(z)\right). \end{aligned}$$

where $\alpha := 1 + \log(z) - \log(2) \in \mathbb{R}$, and where the function ψ_y is defined by

$$(A.3) \quad \psi_y(u) \equiv uy - \frac{1}{2} \sqrt{u} \log(u) + \alpha \sqrt{u} + \frac{1}{2} \left(\mu - \frac{1}{2}\right) \log(u).$$

For $y > 0$ small enough, the saddlepoint equation $\partial_u \psi_y(u) = 0$, or

$$2\mu - 1 + 4uy + 2(\alpha - 1)\sqrt{u} - \sqrt{u} \log(u) = 0,$$

(namely (2.7)) admits a solution $u_y > 0$. This saddlepoint equation can be rewritten as

$$(A.4) \quad y = \frac{\log(u_y)}{4\sqrt{u_y}} + \frac{1 - \alpha}{2\sqrt{u_y}} - \frac{(\mu - 1/2)}{2u_y}.$$

Remark A.1. Note that the saddlepoint equation also reads

$$y = \frac{\log(u_0)}{2\sqrt{2u_0}} - \frac{\rho}{\sqrt{2u_0}} - \frac{4\mu - 2}{4u_0},$$

where $\rho := \log(z/\sqrt{2})$ and $u_0 := 2u_y$, which is reminiscent of that of [18]. In fact, the saddlepoint equation above does not admit a unique solution; in order for the latter to be continuous (as a function of y), one should take the largest solution.

Following [18], we deform the integration contour in (A.1) around the saddlepoint u_y to obtain

$$(A.5) \quad m_y(\mu, z) = \frac{e^{-z}}{2i\pi} \int_{R-i\infty}^{R+i\infty} \Phi(u, y, z) du \sim \frac{\sqrt{z}}{2i\pi} \int_{R-i\infty}^{R+i\infty} e^{\psi_y(u)} du \sim \frac{\sqrt{z}}{2i\pi} \int_{u_y-i\infty}^{u_y+i\infty} e^{\psi_y(u)} du,$$

as y tends to zero. Let λ denote the real integration variable, so that $u = u_y + i\lambda$. Around the saddlepoint ($\lambda = 0$), we have the uniform Taylor series expansions:

$$\begin{aligned} \sqrt{u} &= \sqrt{u_y} + \frac{i\lambda}{2\sqrt{u_y}} + \frac{\lambda^2}{8u_y^{3/2}} + \mathcal{O}\left(\frac{\lambda^3}{u_y^{5/2}}\right), & \log u &= \log u_y + \frac{i\lambda}{u_y} + \frac{\lambda^2}{2u_y^2} + \mathcal{O}\left(\frac{\lambda^3}{u_y^3}\right), \\ \sqrt{u} \log u &= \sqrt{u_y} \log u_y + \frac{(2 + \log(u_y))i\lambda}{2\sqrt{u_y}} + \frac{\log(u_y)\lambda^2}{8u_y^{3/2}} + \mathcal{O}\left(\frac{(1 + \log(u_y))\lambda^3}{u_y^{5/2}}\right), \end{aligned}$$

so that

$$(A.6) \quad \psi_y(u) = u_y y + \sqrt{u_y} \left(\alpha - \frac{\log(u_y)}{2} \right) - \frac{\log(u_y)}{4} + \frac{\mu \log(u_y)}{2} - M_y \lambda^2 + \mathcal{O} \left[\frac{\lambda^3 (1 + \log(u_y))}{u_y^{5/2}} \right]$$

where the coefficients in front of λ cancel out from the saddlepoint equation (2.7), and where

$$M_y := \frac{\log(u_y)}{16u_y^{3/2}} - \frac{\alpha}{8u_y^{3/2}} + \frac{1 - 2\mu}{8u_y^2},$$

as defined in Proposition 2.4. By bootstrapping (see Section A.1.1 for details), the expansion

$$(A.7) \quad u_y = \frac{\log(y)^2}{4y^2} \left[1 - \frac{2 \log \log(1/y)}{\log(y)} + \frac{\log(z^2)}{\log(y)} + o \left(\frac{1}{\log(y)} \right) \right]$$

holds for the saddlepoint as y tends to zero, and implies

$$(A.8) \quad M_y = \frac{y^3}{\log(y)^2} \left[1 + \mathcal{O} \left(\frac{\log |\log(y)|}{\log(y)} \right) \right].$$

Since (A.5) can be rewritten as

$$\begin{aligned} m_y(\mu, z) &\sim \frac{\sqrt{z}}{2i\pi} \int_{u_y - ih}^{u_y + ih} e^{\psi_y(u)} du \\ &\sim \frac{\sqrt{z}}{2\pi} \exp \left[u_y y + \sqrt{u_y} \left(\alpha - \frac{\log(u_y)}{2} \right) - \frac{\log(u_y)}{4} + \frac{\mu \log(u_y)}{2} \right] \int_{-h}^h e^{-M_y \lambda^2} d\lambda, \end{aligned}$$

we need to determine an estimate for the last integral on the compact interval $[-h, h]$. As explained below—where the tail integrals are taken into account—the choice $h := \log(y)^2 / y^{3/2}$ is in fact the right one, and it is then easy to show that, as y tends to zero,

$$\int_{-h}^h e^{-M_y \lambda^2} d\lambda = \frac{\sqrt{\pi} |\log(y)|}{y^{3/2}} + \mathcal{O} \left(e^{-\log(y)^2} y^{-3/2} \right),$$

which then implies

$$\begin{aligned} m_y(\mu, z) &\sim \frac{\sqrt{z}}{2\pi} \exp \left[u_y y + \sqrt{u_y} \left(\alpha - \frac{\log(u_y)}{2} \right) - \frac{\log(u_y)}{4} + \frac{\mu \log(u_y)}{2} \right] \frac{\sqrt{\pi} |\log(y)|}{y^{3/2}} \\ &= \frac{\sqrt{z}}{2\pi} \exp \left[\left(\frac{1}{2} - \mu \right) + \frac{\log(u_y)}{2} \left(\mu - \frac{1}{2} \right) - u_y y + \sqrt{u_y} \right] \frac{\sqrt{\pi} |\log(y)|}{y^{3/2}} \\ &= \frac{\sqrt{z}}{2\sqrt{\pi}} \exp \left(\frac{1}{2} - \mu \right) \frac{|\log(y)|}{y^{3/2}} u_y^{\frac{1}{2}(\mu - \frac{1}{2})} \exp(-u_y y + \sqrt{u_y}) \\ (A.9) \quad &= \frac{z^{\frac{1}{2}} |\log(y)|}{2\sqrt{\pi}} \exp \left\{ -\frac{\log(y)^2}{4y} + \frac{1}{2} \frac{|\log(y)|}{y} + \left(\frac{1}{2} - \mu \right) \left[1 - \frac{1}{2} \log \left(\frac{\log(y)^2}{4y^2} \right) \right] \right\} \left[\frac{1}{y^{\frac{3}{2}}} + \mathcal{O} \left(y^{\frac{3}{2}} \right) \right], \end{aligned}$$

where we used the saddlepoint equation (A.4) in the fourth line.

It now remains to prove that one can indeed neglect the tails of the integration domain, where $\Im(u) = \lambda \geq h$. The analysis of this is similar to that of [18, Section 3], and we only outline here the main arguments. First, specify a choice $h := \log(y)^2 / y^{3/2}$ of integration bounds accounting for the main contribution to the integral $\int_{u_y - i\infty}^{u_y + i\infty} \exp(\psi_y(u)) du$, with ψ_y defined in (A.1) and where u_y denotes the saddlepoint in (A.4). By symmetry, it is clearly sufficient to consider only one side of the tails, and we shall therefore focus on the positive one $\int_{u_y + ih}^{u_y + i\infty} e^{\psi_y(u)} du$. The analysis is then split into studying the inner tail $h \leq \lambda < \exp(\log(1/t)^2/4)$ and the outer tail $\lambda \geq \exp(\log(1/t)^2/4)$. Similarly to [18, Equation (10)], the estimate

$$\int_{u_y + ih}^{u_y + i\infty} e^{\psi_y(u)} du \sim 2 \exp \left\{ u_y t + \frac{1}{8} \log(y)^2 - \exp \left(\frac{\log(y)^2}{8} \right) \right\}$$

prevails for the outer tail. For any real number B , [18, Lemma 1] remains valid for the behaviour of the real part of $\sqrt{u} \log(u) + B\sqrt{u}$ with respect to $|\Im(u)|$, which allows to bound above the inner tail by the value of the integrand at $\lambda = h$ of $-M_y \lambda^2|_{\lambda=h} \sim -\frac{1}{2} \log(y)^2$ multiplied by the length of the integration path, which is of order $e^{\log(1/t)^2/4}$; the relative error is therefore of order $\exp(-\frac{1}{4} \log(y)^2 + o(\log(y)^2))$.

The final part of the error analysis in the expansion (A.9) follows from analogous estimates to [18, Table 1], and the total (both tails) error resulting from the completion to Gaussian integral

$$\frac{2}{\sqrt{2M_y}} \int_h^\infty \frac{\exp\left(-\frac{1}{2}\omega^2\right)}{\sqrt{2M_y}} d\omega \sim \frac{2}{\sqrt{2M_y}} \frac{\exp\left(-\frac{1}{2}\omega^2\right)}{\omega} \Big|_{\omega=h\sqrt{M_y}} = \exp\left(-\frac{1}{2} \log(t)^2 + o(\log(t))\right).$$

The error $\mathcal{O}(\lambda^3/u_y^{5/2})$ from the local expansion (A.6) for ψ_y is of order

$$(A.10) \quad \mathcal{O}(\log(y)^2 \sqrt{y}),$$

which is immediate from bootstrapping (A.8), (A.7) for M_y and u_y , and from the choice of h ,

$$\frac{\lambda^3}{u_y^{5/2}} \leq C \frac{\log(u_y)}{u_y^{3/2}} \frac{1}{u_y} \frac{\log(y)^2}{y^{3/2}} \sim C \log(y)^2 \sqrt{y}.$$

Hence the total relative error is dominated by the error (A.10) from the local expansion if M_y is not expanded, and by the relative error (A.8) of M_y if one consider its bootstrapping expansion.

A.1.1. *Justification of the expansion for M_y .* We define

$$\widetilde{M}_y := \frac{\log(u_y)}{16u_y^{3/2}} - \frac{\alpha}{8u_y^{3/2}}.$$

The term $(1-2\mu)/(8u_y^2)$ in the definition (2.8) of M_y is of higher order, so that we can work with the simpler expression $\widetilde{M}_y$ in the bootstrapping expansion and the error analysis instead of M_y . With $\alpha = \rho + 1 - \frac{1}{2} \log(2)$ and $\tilde{u} \equiv u_y/2$, the approximation of the saddlepoint simplifies to

$$M_y = \frac{\log(u_y)}{16u_y^{3/2}} - \frac{\alpha}{8u_y^{3/2}} = \frac{\log(u_y)}{16u_y^{3/2}} - \frac{\rho+1}{8u_y^{3/2}} + \frac{\log(2)}{16u_y^{3/2}} = \frac{1}{4} \left(\frac{\sqrt{2} \log(\tilde{u})}{16\tilde{u}^{3/2}} - \frac{\sqrt{2}(\rho+1-\log(2))}{8\tilde{u}^{3/2}} \right).$$

Thus M_y is up to constants of the same form as [18, Equation (12)]. By bootstrapping,

$$M_y \sim \frac{y^3}{\log(y)^2} \left[1 + \mathcal{O}\left(\frac{\log|\log(y)|}{\log(y)}\right) \right].$$

Indeed, the saddlepoint equation (2.7)

$$y = \frac{\log(2u_y)}{2\sqrt{2(2u_y)}} - \frac{\rho}{\sqrt{2(2u_y)}} - \frac{4\mu-2}{4(2u_y)},$$

when setting $c(u) \equiv \left(\frac{\log(\sqrt{u})}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u}}\right)$, $\rho = \log\left(\frac{z}{\sqrt{2}}\right)$, $k := 4\mu - 2$ and $u_0 = 2u_y$ becomes $\sqrt{u_0} \equiv y^{-1}c(u_0)$, where $\log(\sqrt{u_0}) = \log(c(u_0) - \log(y))$. Hence, bootstrapping as in [18] yields

$$\begin{aligned} u_0 &= \frac{1}{y^2} \left(\frac{\log(1/y)}{\sqrt{2}} + \frac{\log(c(u_0))}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u_0}} \right)^2 \\ &= \frac{1}{y^2} \left(\left(\frac{\log(1/y)}{\sqrt{2}} \right)^2 + 2 \left(\frac{\log(1/y)}{\sqrt{2}} \right) \left(\frac{\log(c(u_0))}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u_0}} \right) + \left(\frac{\log(c(u_0))}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u_0}} \right)^2 \right) \\ &= \frac{(\log(1/y))^2}{2y^2} \left[1 + \left(\frac{2\sqrt{2}}{\log(1/y)} \right) \left(\frac{\log(c(u_0))}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u_0}} \right) + \frac{2}{(\log(1/y))^2} \left(\frac{\log(c(u_0))}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u_0}} \right)^2 \right]. \end{aligned}$$

Now expanding around $\log(1/y)$,

$$\frac{\log(c(u_0))}{\sqrt{2}} \sim \frac{\log(\log(1/y))}{\sqrt{2}} - \frac{\log(2)}{2\sqrt{2}} + \frac{\log\left(c(u_0) - \rho + \frac{k}{2\sqrt{2}u_0}\right)}{\sqrt{2} \log(1/y)},$$

and using the fact that both

$$-\frac{2\sqrt{2}}{\log(y)} \left(\frac{k}{4\sqrt{u}} - \frac{\log(c(u) - \rho + \frac{k}{2\sqrt{2}u_0})}{\sqrt{2}\log(y)} \right) \quad \text{and} \quad \frac{2}{\log(y)^2} \left(\frac{\log(c(u_0)) - \rho}{\sqrt{2}} + \frac{k}{4\sqrt{u_0}} \right)^2$$

are of order $o(1/\log(y))$, we obtain, by collecting terms,

$$2u_y = \frac{\log(y)^2}{2y^2} \left(1 - \frac{2\log(-\log(y))}{\log(y)} + \frac{2\rho + \log(2)}{\log(y)} + o\left(\frac{1}{\log(y)}\right) \right).$$

Similarly,

$$u_0^{3/2} = \frac{1}{y^3} \left[\frac{-\log(y)}{\sqrt{2}} + \frac{\log(c(u))}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u}} \right]^3 \sim \frac{-\log(y)^3}{2\sqrt{2}y^3} \left[1 - \frac{\log(-\log(y))}{\log(y)} + \frac{2\rho + \log(2)}{2\log(y)} + o\left(\frac{1}{\log(y)}\right) \right]^3,$$

hence $u_y^{3/2} \sim (\log(1/y))^2/(8y^2)$; further,

$$\log(u_0) = -2(\log(y) - \log(c(u))) \sim -2\log(y) + 2\log(-\log(y)) - \log(2) - \frac{2\log(c(u) - \rho + \frac{k}{2\sqrt{2}u})}{\log(y)},$$

so that, by bootstrapping we also recover the form of [18, Equation (13)], at $\tilde{u} = \frac{1}{2}u_y$:

$$M_y = \frac{1}{4} \left[\frac{\sqrt{2}\log(\tilde{u})}{16\tilde{u}^{3/2}} - \frac{\sqrt{2}(\rho + 1 - \log(2))}{8\tilde{u}^{3/2}} \right] = \frac{y^3}{2\log(y)^2} \left[1 + \mathcal{O}\left(\frac{\log(-\log(y))}{\log(y)}\right) \right].$$

REFERENCES

- [1] M. Abramowitz, I. A. Stegun. Handbook of mathematical functions: with formulas, graphs, and mathematical tables. Dover Publications, 1964.
- [2] J. Andreasen and B. Høge. ZABR – expansions for the masses. Preprint, SSRN//1980726, 2011.
- [3] A. Antonov and M. Spector. Advanced analytics for the SABR model. Preprint, SSRN//2026350, 2012.
- [4] A. Antonov, M. Konikov and M. Spector. The free boundary SABR: natural extension to negative rates. *Risk*, September 2015.
- [5] A. Antonov, M. Konikov and M. Spector. Mixing SABR models for negative rates. SSRN//2653682, 2015.
- [6] A. Antonov, M. Konikov and M. Spector. SABR spreads its wings. *Risk*, August 2013.
- [7] P. Balland and Q. Tran. SABR Goes Normal. *Risk*, June issue: 76-81, 2013.
- [8] P. Barrieu, A. Rouault, and M. Yor. A study of the Hartman-Watson distribution motivated by numerical problems related to Asian option pricing. *Journal of Applied Probability*, 41: 1049-1058, 2004.
- [9] S. Benaim, P. Friz. Regular variation and smile asymptotics. *Mathematical Finance*, 19: 1-12, 2009.
- [10] A. N. Borodin, P. Salminen. Handbook of Brownian motion - Facts and Formulae. Birkhäuser, 2nd Ed., 1996.
- [11] B. Chen, C. W. Oosterlee, and H. van der Weide. A low-bias simulation scheme for the SABR stochastic volatility model. *International Journal of Theoretical and Applied Finance*, 15(2), 2012.
- [12] S. De Marco, C. Hillairet, and A. Jacquier. Shapes of implied volatility with positive mass at zero. *SIAM Journal on Financial Mathematics*, 8(1): 709-737, 2017.
- [13] L. Döring, B. Horvath, and J. Teichmann. Functional analytic (ir-)regularity properties of SABR-type processes. *International Journal of Theoretical and Applied Finance*, 20 (3): 1-48, 2017.
- [14] D. Dufresne. The integral of geometric Brownian motion. *Advances in Applied Probability*, 33: 223-241, 2001.
- [15] M. Forde, A. Pogudin. The large-maturity smile for the SABR and CEV-Heston models. *International Journal of Theoretical and Applied Finance*, 16(8), 2013.
- [16] J. Gatheral. The volatility surface: a practitioner's guide. Wiley, 2006.
- [17] J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1): 59-71, 2014.
- [18] S. Gerhold. The Hartman-Watson distribution revisited: asymptotics for pricing Asian options. *Journal of Applied Probability*, 48(3): 892-899, 2011.
- [19] I. S. Gradshteyn and I. M. Ryzhik. Table of integrals, series, and products, 6th Edition. Academic Press, 2000.
- [20] L. A. Grzelak and C. W. Oosterlee. From arbitrage to arbitrage-free implied volatilities. *Journal of Computational Finance*, 20(3): 31-49, 2016.
- [21] A. Gulisashvili. Left-wing asymptotics of the implied volatility in the presence of atoms. *International Journal of Theoretical and Applied Finance*, 18(2), 2015.
- [22] A. Gulisashvili. Analytically tractable stochastic stock price models. Springer, 2012.
- [23] A. Gulisashvili. Asymptotic formulas with error estimates for call pricing functions and the implied volatility at extreme strikes. *SIAM Journal on Financial Mathematics*, 1: 609-641, 2010.
- [24] A. Gulisashvili, B. Horvath and A. Jacquier. On the probability of hitting the boundary for Brownian motions on the SABR plane. *Electronic Communications in Probability*, 21(75): 1-13, 2016.

- [25] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward. Managing smile risk. *Wilmott Magazine*, September issue: 84-108, 2002.
- [26] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward. Arbitrage-free SABR. *Wilmott Magazine*, January issue: 60-75, 2014.
- [27] P. Hagan, A. Lesniewski, and D. Woodward. Probability distribution in the SABR model of stochastic volatility. *Large Deviations and Asymptotic Methods in Finance (Editors: P. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, J. Teichmann)*, Springer Proceedings in Mathematics and Statistics, **110**, 2015.
- [28] D. Hobson. Comparison results for stochastic volatility models via coupling. *Finance and Stochastics*, **14**: 129-152, 2010.
- [29] O. Islah. Solving SABR in exact form and unifying it with LIBOR market model. SSRN//1489428, 2009.
- [30] J. Jacod and A.N. Shiryaev. Limit theorems for stochastic processes, Second Edition. Springer Berlin, 2003.
- [31] M. Jeanblanc, M. Yor, and M. Chesney. Mathematical methods for financial markets. Springer Finance, 2009.
- [32] B. Jourdain. Loss of martingality in asset price models with lognormal stochastic volatility. *International Journal of Theoretical and Applied Finance*, **13**: 767-787, 2004.
- [33] R. W. Lee. The moment formula for implied volatility at extreme strikes. *Math Finance*, **14**: 469-480, 2004.
- [34] A. Leita, L. A. Grzelak, and C. W. Oosterlee. On a one time-step Monte Carlo simulation approach of the SABR model: application to European options. *Applied Math. and Computation*, **293**(15): 461-479, 2017.
- [35] M. Lorig, S. Pagliarani, A. Pascucci. Explicit implied volatilities for multifactor local-stochastic volatility models To appear in Math. Finance, 2016.
- [36] H. Matsumoto and M. Yor. Exponential functionals of Brownian motion, I: Probability laws at fixed time. *Probability Surveys*, **2**: 312-347, 2005.
- [37] H. Matsumoto and M. Yor. Exponential functionals of Brownian motion, II: Some related diffusion processes. *Probability Surveys*, **2**: 348-384, 2005.
- [38] P. D. Miller. Applied Asymptotic Analysis. Graduate Studies in Mathematics, volume 75, American Mathematical Society, 2006.
- [39] J. Oblój. Fine-tune your Smile: Correction to Hagan et al. *Wilmott Magazine*, May issue, 2008.
- [40] S. Pagliarani and A. Pascucci. Asymptotic expansions for degenerate parabolic equations. *C. R. Acad. Sci. Paris, Ser. I*(352): 1011-1016, 2014.
- [41] S. Pagliarani and A. Pascucci. The exact Taylor Formula of the Implied Volatility. *Finance and Stochastics*, **21**(3): 661-718, 2017.
- [42] L. Paulot. Asymptotic implied volatility at the second order with application to the SABR model. *Large deviations and asymptotic methods in finance (Editors: P. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, J. Teichmann)*, Springer Proceedings in Mathematics and Statistics, **110**, 2015.
- [43] R. Rebonato. A simple approximation for the no-arbitrage drifts for LMM-SABR-family interest-rate models. *Journal of Computational Finance*, **19**(1), 2015.
- [44] D. Revuz and M. Yor. Continuous martingales and Brownian motion. Springer, Berlin, 2004.
- [45] A. Veraart and M. Winkel. Time change. *Encycl. of Quant. Fin.* (Editor R. Cont), Wiley, **IV**: 1812-1816, 2010.
- [46] M. Yor. On some exponential functionals of Brownian motion. *Adv. Appl. Prob.*, **24**: 509-531, 1992.

DEPARTMENT OF MATHEMATICS, OHIO UNIVERSITY
 E-mail address: `gulisash@ohio.edu`

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON
 E-mail address: `b.horvath@imperial.ac.uk`

DEPARTMENT OF MATHEMATICS, IMPERIAL COLLEGE LONDON AND BARUCH COLLEGE, CUNY
 E-mail address: `a.jacquier@imperial.ac.uk`