MASS AT ZERO AND SMALL-STRIKE IMPLIED VOLATILITY EXPANSION IN THE SABR MODEL

ARCHIL GULISASHVILI, BLANKA HORVATH, AND ANTOINE JACQUIER

Abstract. We study the probability mass at the origin in the SABR stochastic volatility model, and derive several tractable expressions for it, in particular when time becomes small or large. In the uncorrelated case, tedious saddlepoint expansions allow for (semi) closed-form asymptotic formulae. As an application–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 currently used implied volatility expansions. In particular we discuss how much those expansions become erroneous.

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 [3, 4, 6, 34]. It is defined by the pair of coupled stochastic differential equations

$$
\begin{aligned}
\mathrm{d}X_t &= Y_t X_t^\beta \mathrm{d}W_t, \qquad & X_0 = x_0 > 0,\\
\mathrm{d}Y_t &= \nu Y_t \mathrm{d}Z_t, \qquad & Y_0 = y_0 > 0,\\
\mathrm{d}\langle Z, W\rangle_t &= \rho \mathrm{d}t,
\end{aligned}
\tag{1.1}
$$

where $\nu > 0$, $\rho \in (-1, 1)$, $\beta \in (0, 1)$, and W and Z are two correlated Brownian motions living on a given 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 process X in (1.1), therefore exhibiting arbitrage.

There exist several refinements to this asymptotic formula: in [41] Obłój 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 have been obtained in [5, 16, 29] by applying time-change techniques, and the correlated case can be seen as an approximation of the uncorrelated case (see for example [4, 5] using a projection to a mimicking model). However, it seems that these refinements have not fully resolved the arbitrage issue near the origin. This problem of negative density in low interest-rate environments has been directly addressed by Hagan et. al [26], Balland and Tran [6], and Andreasen and Huge [3], who propose modifications of the original SABR model or of the base model for the implied volatility expansion. The original asymptotic formula typically loses accuracy for long-dated derivatives, when the CEV exponent (the parameter β) is close to zero, or when the volatility of volatility ν is large. The parameter β determines the backbone and

Date: February 12, 2015.

2010 *Mathematics Subject Classification.* 58J37, 60H30, 58J65.

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

The authors would like to Rama Cont and Josef Teichmann for organising the (still-running) successful series of ETH-Imperial College workshops, where this project initiated. They are also indebted to Andrei Borodin for stimulating discussions regarding densities of functionals of Brownian motion. BH would like to thank Leif Döring and Leonid Mytnik for stimulating discussions on time change techniques. AJ acknowledges financial support from the EPSRC First Grant EP/M008436/1. Part of this work has been done on the collaborative platform Zanadu (www.zanadu.io).

governs the dynamics of the smile, and small values are usually chosen when the asymptotic formula fails, namely on markets where the forward rate is close to zero and for long-dated options [6, 25]. It therefore comes as no surprise that the Hagan formula–which is an asymptotic expansion for small values of ν^2T–breaks down in such a setting.

Approximations of the implied volatility in terms of asymptotic expansions are available, not only for small and large maturities, but also for extreme strikes. Roger Lee's celebrated Moment Formula [35] relates the behaviour of the implied volatility $I_T(K)$ for small strikes K to the price of a Put option on X. This model-independent result was subsequently refined by Benaim and Friz [8] and Gulisashvili [24]. In [12, 22] De Marco et. al. and Gulisashvili independently showed that when the underlying distribution has an atom at zero (which is often the case in the SABR model), the small-strike behaviour of the implied volatility is solely determined by this mass, irrespective of the distribution of the process on $(0,\infty)$. In order to understand the small-strike behaviour of the SABR smile, we study the probability $\mathbb{P}(X_T = 0)$–and provide tractable formulae and asymptotic approximations thereof–that the forward price (1.1) lies at the origin. Since much of the popularity of the SABR model is due to the tractability of its asymptotic formula, it is indeed desirable to preserve the latter when the mass at zero is present.

The paper is organised as follows: 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 [5, 11, 16, 29] to determine the exact distribution of the absolutely continuous part of the distribution. 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 Section 2.3, we derive asymptotic expansions of the atom at the origin in the short time and long time cases.

In Section 3 we study a drift extended version of (1.1), which we refer to as the Brownian motion on the SABR plane (see (3.1)), and study its dynamics under different configurations of the SABR parameters. Of particular interest are the zero correlation case, which we relate to Section 2, and the case $\beta = 0$ (and general ρ), which is prevalently chosen on markets when the current asymptotic formula breaks down. Besides, when $\beta = 0$, the original SABR dynamics (1.1) coincide with those of the Brownian motion on the SABR plane, and we propose several space transformations to translate the properties of one parameter configuration to another, thus deriving an explicit formula for the mass at zero for large times when $\beta = 0$ in the correlated case.

In Section 4, we use the results of the previous sections to determine the left wing of the SABR implied volatility. Using the formulae provided in [12, 22], we highlight the fact that some of the widely used expansions exhibit arbitrage in the left wing, and we propose a way to regularise them in this arbitrageable region.

Preliminaries and Notations: the process X in (1.1) is a martingale [33, Remark 2]. If we consider X on the state space $[0,\infty)$, then the origin, which can be attained, has to be absorbing (see [31, Chapter III, Lemma 3.6]). For a given real-valued stochastic process X (with continuous paths) and a real number z, we define by $\tau_z^X := \inf\{t \geq 0 : X_t = z\}$ the first hitting time of X at level z. For convenience, we shall use the (now fairly standard) notation $\overline{\rho} := \sqrt{1-\rho^2}$. For two functions f and g, we shall write $f(z) \sim g(z)$ as z tends to zero whenever $\lim\limits_{z\to 0} f(z)/g(z) = 1$.

2. Mass at zero in the uncorrelated SABR model

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 $\widehat{X}$ satisfies the CEV stochastic differential equation

$$\mathrm{d}\widehat{X}_t = \widehat{Y}_t \widehat{X}_t^\beta \mathrm{d}W_t, \quad \widehat{X}_0 = x_0,$$

where $\widehat{Y}$ 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 $\mathrm{d}\widetilde{X}_t = \widetilde{X}_t^\beta \mathrm{d}W_t$ starting from x_0, and set

$$\widehat{G}_t := \frac{\widehat{X}_t^{2(1-\beta)}}{(1-\beta)^2} \qquad \text{and} \qquad \widetilde{G}_t := \frac{\widetilde{X}_t^{2(1-\beta)}}{(1-\beta)^2}.$$

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

$$\mathrm{d}Z_t = \frac{1-2\beta}{1-\beta}\mathrm{d}t + 2\sqrt{|Z_t|}\mathrm{d}W_t, \qquad Z_0 = \frac{x_0^{2(1-\beta)}}{(1-\beta)^2}.$$

By Itô's formula, the process $\widetilde{G}$ solves the same SDE, so that $Z = \widetilde{G}$, and therefore $\widehat{X}_t = \widetilde{X}_{\int_0^t \widehat{Y}_s^2 \mathrm{d}s}$, for all $t \geq 0$. It follows that X can be obtained from $\widetilde{X}$ using the stochastic time change

$$t \mapsto \int_0^t Y_s^2 \mathrm{d}s, \tag{2.1}$$

namely $X_t = \widetilde{X}_{\int_0^t Y_s^2 \mathrm{d}s}$. Since this time change process is independent of $\widetilde{X}$, one can decompose the mass at zero of the SABR model into the mass of the CEV component at zero and the density of the time change:

$$\mathbb{P}\left(X_t = 0\right) = \int_0^\infty \mathbb{P}\left(\widetilde{X}_r = 0\right) \mathbb{P}\left(\int_0^t Y_s^2 \mathrm{d}s \in \mathrm{d}r\right) \mathrm{d}r, \tag{2.2}$$

where the mass at zero in the CEV model is given by (see [12] or [32])

$$\mathbb{P}\left(\widetilde{X}_r = 0\right) = 1 - \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right), \tag{2.3}$$

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

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 (2.3). 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 (2.3) 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 [31, 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

$$\mathbb{P}\left(\int_0^t Y_s^2 \mathrm{d}s \in \mathrm{d}r\right) = \mathbb{P}\left(\int_0^t \mathrm{e}^{2\nu Z_s^{(-\nu/2)}} \mathrm{d}s \in \mathrm{d}\widetilde{r}\right), \tag{2.4}$$

where $\widetilde{r} := r/y_0^2$ and $Z_s^{(-\nu/2)} := Z_s - \frac{1}{2}\nu s$. In [10, Formula 1.10.4, page 264], the density of the above functional is given by (see also Section 2.4.3 below for details)

$$\mathbb{P}\left(\int_0^t \mathrm{e}^{2\nu Z_s^{(-\nu/2)}} \mathrm{d}s \in \mathrm{d}\widetilde{r}\right) = \frac{2^{1/4}\sqrt{\nu}}{\widetilde{r}^{3/4}} \exp\left(-\frac{\nu^2 t}{8} - \frac{1}{4\nu^2 \widetilde{r}}\right) m_{2\nu^2 t}\left(-\frac{3}{4}, \frac{1}{4\nu^2\widetilde{r}}\right) \mathrm{d}\widetilde{r}, \tag{2.5}$$

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

$$m_y(\mu, z) \equiv \frac{8z^{3/2}\Gamma(\mu + \frac{3}{2})\mathrm{e}^{\frac{\pi^2}{4y}}}{\pi\sqrt{2\pi y}} \int_0^\infty \mathrm{e}^{-z\cosh(2u) - \frac{1}{y}u^2} \mathrm{M}\left(-\mu, \frac{3}{2}, 2z\sinh(u)^2\right) \sinh(2u) \sin\left(\frac{\pi u}{y}\right) \mathrm{d}u, \tag{2.6}$$

and where the Kummer function M reads

$$\mathrm{M}(a, b, x) \equiv 1 + \sum_{k=1}^\infty \frac{a(a+1)\dots(a+k-1)x^k}{b(b+1)\dots(b+k-1)k!}. \tag{2.7}$$

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 over 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 [23, Section 4.6]. These numerical issues are discussed in [7], and Gerhold [19] used saddlepoint methods to provide short-time estimates. Because of the time change and the complexity of the Kummer function (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 [19], to provide small-time asymptotic estimates for the density of the time change. In Section 2.4 below, we provide several alternative representations for this density, and relate them to the existing literature. We recall the mapping $y = 2\nu^2 t$, and we shall alternate between the two notations without ambiguity in order to simplify some formulations.

Remark 2.2. Let $\omega := 1/y$. The function m has the form

$$m_\omega(\cdot) = c_\omega \int_0^\infty \mathrm{e}^{-u^2\omega} f_\omega(u)\mathrm{d}u,$$

for some c_ω and f_ω. One might be tempted to use a standard Laplace method here to determine the asymptotic behaviour as ω 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_ω–appearing as coefficients of the expansion–are null, and hence the method does not apply.

We now formulate one of the main results of the paper, which characterises the small-time asymptotic 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

$$2\mu - 1 + 4uy + 2\log(z/2)\sqrt{u} - \sqrt{u}\log(u) = 0, \tag{2.8}$$

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

$$M_y := \frac{\log(u_y)}{16u_y^{3/2}} - \frac{\alpha}{8u_y^{3/2}} + \frac{1-2\mu}{8u_y^2}. \tag{2.9}$$

Theorem 2.3. *As t tends to zero, the following asymptotic formula holds for the mass of the atom at zero in the uncorrelated SABR model:*

$$\mathbb{P}(X_t = 0) \sim \frac{y_0^{3/2}\mathrm{e}^{5/4}}{2^{7/4}\sqrt{\nu\pi}} \exp\left(-\frac{\nu^2 t}{8}\right) \int_0^\infty \exp\left\{\frac{\log(u_y)}{2}\left(\mu - \frac{1}{2}\right) - u_y y + \sqrt{u_y}\right\} \frac{g(r)}{\sqrt{M_y}}\mathrm{d}r,$$

where

$$g(r) \equiv \mathbb{P}\left(\widetilde{X}_r = 0\right) \frac{1}{r^{5/4}} \exp\left(-\frac{y_0^2}{4\nu^2 r}\right).$$

Theorem 2.3 follows from (2.2) and the following assertion.

Proposition 2.4. *As y (equivalently t) tends to zero, we have (recall that $y = 2\nu^2 t$)*

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

The technical part of the proof relies on the following proposition, proved in Appendix A.1.

Proposition 2.5. *As y tends to zero, the function m_y in (2.6) satisfies*

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

Remark 2.6. The proof of the proposition uses saddlepoint analysis. The saddlepoint u_y is the solution to the equation (2.8), 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

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

but numerical computations show that this estimate is not very accurate.

2.3. **Large-time asymptotics.** We now concentrate on the large-time behaviour of the mass at zero in the uncorrelated SABR model. From [10, Formula 1.8.4, page 612], the formula

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

holds, so that the decomposition (2.2) together with (2.4) implies that the mass at zero reads (recall that $\widetilde{r} = r/y_0^2$)

$$\begin{aligned} \mathbb{P}_\infty := \lim_{t\uparrow\infty} \mathbb{P}(X_t = 0) &= \frac{y_0}{\nu\sqrt{2\pi}} \int_0^\infty \left[1 - \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right)\right] r^{-3/2} \exp\left(-\frac{y_0^2}{2\nu^2 r}\right) \mathrm{d}r \\ &= 1 - \frac{y_0}{\nu\sqrt{2\pi}} \int_0^\infty \Gamma\left(\frac{1}{2(1-\beta)}, \frac{x_0^{2(1-\beta)}}{2r(\beta-1)^2}\right) r^{-3/2} \exp\left(-\frac{y_0^2}{2\nu^2 r}\right) \mathrm{d}r. \end{aligned} \tag{2.10}$$

In the case where $\beta = 0 (= \rho)$, the SABR model (1.1) reduces to a Brownian motion on the hyperbolic plane (up to a deterministic time change, see Section 3.1), and a simple computation shows that (2.10) simplifies to

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

When $\beta \neq 0$, the integral in (2.10) 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 nth-order approximation

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

Note in particular that

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

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 2.7 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

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

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

$$\left|\exp\left\{-\frac{y_0^2}{2\nu^2 r}\right\} - \sum_{k=0}^{n}(-1)^k \frac{1}{n!}\left(\frac{y_0^2}{2\nu^2 r}\right)^k\right| \leq \frac{1}{(n+1)!}\left(\frac{y_0^2}{2\nu^2 r}\right)^{n+1}. \tag{2.11}$$

For any $n \geq 0$, set

$$b_n := \frac{2y_0(1-\beta)}{\Gamma\left(\frac{1}{2(1-\beta)}\right)\nu\sqrt{\pi}x_0^{1-\beta}}\left(\frac{y_0^2(\beta-1)^2}{\nu^2 x_0^{2(1-\beta)}}\right)^n \frac{\Gamma\left(n+1+\frac{\beta}{2-2\beta}\right)}{n!(1+2n)}, \tag{2.12}$$

so that from (2.11), (2.12), and the definitions of $\mathbb{P}_\infty$ and $\mathbb{P}_\infty^{(n)}$, it follows that

$$\mathbb{P}_\infty^{(n)} = \sum_{k=0}^{n}(-1)^k b_k, \quad n \geq 0, \tag{2.13}$$

and

$$\left|\mathbb{P}_\infty - \mathbb{P}_\infty^{(n)}\right| \leq b_{n+1}, \quad n \geq 0. \tag{2.14}$$

Theorem 2.7. *The following statements hold for the sequence* $\mathbb{P}^{(n)}$ *in* (2.13)*:*

(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,*

$$\mathbb{P}_\infty = \mathbb{P}_\infty^{(n)} + \mathcal{O}\left(n^{-1+\frac{\beta}{2-2\beta}}\exp\left(-n\log\left(\frac{\nu^2 x_0^{2(1-\beta)}}{y_0^2(\beta-1)^2}\right)\right)\right), \quad \text{as } n \text{ tends to infinity.} \tag{2.15}$$

Proof. From (2.12), Stirling's formula for the Gamma function yields

$$b_k \sim \frac{y_0(1-\beta)}{\Gamma\left(\frac{1}{2(1-\beta)}\right)\nu\sqrt{\pi}x_0^{1-\beta}} k^{-1+\frac{\beta}{2-2\beta}}\left(\frac{y_0^2(\beta-1)^2}{\nu^2 x_0^{2(1-\beta)}}\right)^k, \tag{2.16}$$

as k tends to infinity. From (2.13), if the conditions of Theorem 2.7(i) hold, then the general term of the series $\sum_{k=0}^{\infty}(-1)^k b_k$ does not tend to zero, and hence the sequence $(\mathbb{P}_\infty^{(n)})_{n\geq 0}$ diverges. On the other hand, if the conditions of Theorem 2.7(ii) hold, then (2.14) and (2.16) imply (2.15), which completes the proof of Theorem 2.7. □

Remark 2.8. For practical purposes, depending on whether the conditions for convergence of the sequence $(\mathbb{P}_\infty^{(n)})_{n\geq 0}$ in Theorem 2.7, it may or may not be useful to use directly the integral form (2.10). Consider the following values: $(y_0, \nu, \beta, x_0) = (0.1, 1.0, 0.2, 0.2)$, for which convergence is ensured. The exact mass at zero in this case is $\mathbb{P}_\infty = 20.833\%$. Using Theorem 2.7, the table below computes the error using the sequence $(\mathbb{P}_\infty^{(n)})_{n\geq 0}$:

	$n=0$	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
$\lvert\mathbb{P} - \mathbb{P}_\infty^{(n)}\rvert$	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

and the table below computes the integral (2.10) 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

2.3.1. *Large-time numerics.* We provide below some numerics of the large-time mass at zero derived in (2.10). In particular, we observe the influence of the parameter β (Figure 1) as well as that of the starting point x_0 (Figure 1) 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 4 below.

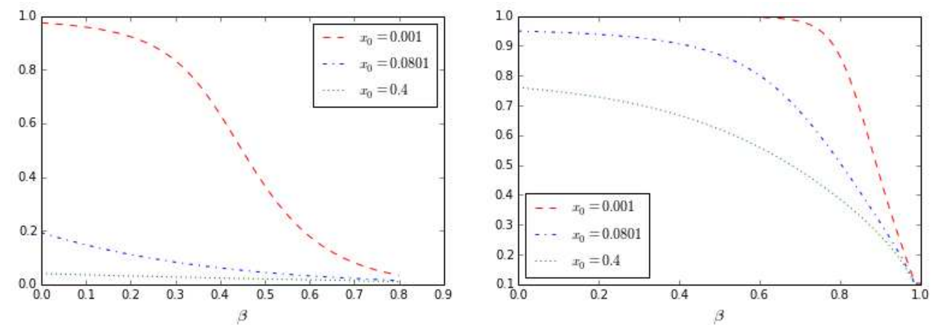

FIGURE 1. Influence of β on the large-time mass at zero in the uncorrelated SABR model with $(y_0, \nu) = (0.015, 0.6)$ (left) and $(y_0, \nu) = (0.1, 1)$ (right).

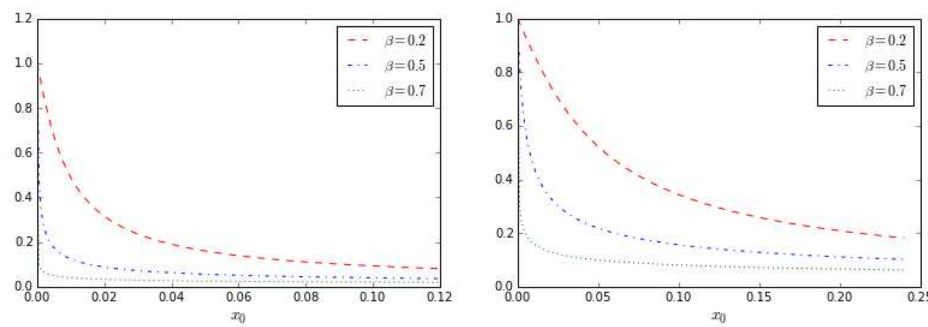

FIGURE 2. Influence of the initial value x_0 on the large-time mass at zero with $(y_0, \nu) = (0.015, 0.6)$ (left) and $(y_0, \nu) = (0.1, 1)$ (right).

2.4. **Representations of the density of the integrated variance.** We are interested here in computing tractable expressions for the density of the random variable $\int_0^t Y_s^2 \mathrm{d}s$ in the SABR model (1.1). In order to set the notations, consider the following drifted version of the latter:

$$
\begin{aligned}
\mathrm{d}X_t &= Y_t X_t^{\beta} \mathrm{d}W_t, & X_0 &= x_0 \in \mathbb{R}, \\
\mathrm{d}Y_t &= \left(\mu + \frac{1}{2}\right) Y_t \mathrm{d}t + Y_t \mathrm{d}Z_t, & Y_0 &= y_0 > 0,
\end{aligned} \tag{2.17}
$$

for some $\mu \in \mathbb{R}$, and we shall, unless otherwise stated, assume that $y_0 = 1$. Let further $\mathbf{Y}^{(\mu)}$ denote the integrated variance process

$$
\mathbf{Y}_t^{(\mu)} := \int_0^t Y_s^2 \mathrm{d}s = y_0 \int_0^t \mathrm{e}^{2Z_s^{(\mu)}} \mathrm{d}s, \tag{2.18}
$$

where we recall that $Z_s^{(\mu)} := Z_s + \mu s$. Finally, for any $t \geq 0$, we shall denote by $\varphi_t^{(\mu)}$ and $\lambda_t^{(\mu)}$ the probability density functions, on $(0, \infty)$, of the random variables $\mathbf{Y}_t^{(\mu)}$ and $1/(2\mathbf{Y}_t^{(\mu)})$. We first start with a 'toy' version, for which computations are simpler, before extending them to the standard case.

2.4.1. *The 'toy' SABR model.*

Definition 2.9. The 'toy' SABR model is the unique strong solution to (2.17) with $\mu = 0$, $y_0 = 1$.

Define the function $\Psi : (0, \infty) \times \mathbb{R} \to \mathbb{R}$ by

$$\Psi(t, w) \equiv \frac{1}{\sqrt{(1+2w)t}} \exp\left(-\frac{\left(\log\left[(\sqrt{2w})^{1/2} + (\sqrt{2w+1})^{1/2}\right]\right)^2}{2t}\right). \tag{2.19}$$

The next lemma derives an expression for the density $\varphi_t^{(0)}$ of the random variable $\mathbf{Y}_t^{(0)}$.

Lemma 2.10. *For all $t, x, R > 0$,*

$$\varphi_t^{(0)}\left(\frac{1}{x}\right) = \frac{x^{3/2}}{2\mathrm{i}\pi} \int_{R-\mathrm{i}\infty}^{R+\mathrm{i}\infty} \Psi(t, z)\mathrm{e}^{xz}\mathrm{d}z = \frac{x^{3/2}}{2\pi}\mathrm{e}^{Rx} \int_{-\infty}^{\infty} \Psi(t, R + \mathrm{i}s)\mathrm{e}^{\mathrm{i}xs}\mathrm{d}s.$$

Proof. Recall Bougerol's identity [23, Equation 4.9]: let Z and $Z^\perp$ be two independent Brownian motions and let $\mathbf{Y}^{(0)}$ be the process defined in (2.18). Then $\sinh(Z_t)$ and $Z^\perp_{\mathbf{Y}_t^{(0)}}$ have the same law for each $t \geq 0$. It follows that, for any $z \in \mathbb{R}$,

$$\begin{aligned}\frac{1}{\sqrt{2\pi t(1+z^2)}} \exp\left(-\frac{(\operatorname{arcsinh}(z))^2}{2t}\right) &= \frac{1}{\sqrt{2\pi}} \int_0^\infty y^{-1/2} \exp\left(-\frac{z^2}{2y}\right) \varphi_t^{(0)}(y)\mathrm{d}y \\ &= \int_0^\infty u^{-3/2} \varphi_t^{(0)}\left(\frac{1}{u}\right) \exp\left(-\frac{z^2 u}{2}\right) \mathrm{d}u,\end{aligned}$$

with the substitution $y = u^{-1}$. Next, using $w = \frac{1}{2}z^2$ and the identity $\operatorname{arcsinh}(a) = \log(\sqrt{a} + \sqrt{a+1})$, we obtain, for all $t > 0$ and $w \in \mathbb{C}_+$,

$$\int_0^\infty u^{-3/2} \varphi_t^{(0)}\left(\frac{1}{u}\right) \mathrm{e}^{-wu}\mathrm{d}u = \Psi(t, w), \tag{2.20}$$

where the function Ψ is defined in (2.19). The lemma then follows from applying the Mellin inverse formula to the Laplace transform (2.20). Note that, from (2.19), the function $\Psi(t, \cdot)$ clearly admits an analytic extension to $\mathbb{C}_+$ for every $t > 0$. Therefore, we can extend formula (2.20) to $\mathbb{C}_+$, where we use the principal branches of the functions $w \mapsto \sqrt{w}$ and $w \mapsto \log(x)$. □

2.4.2. *Density of the integrated variance.* We now return to the standard SABR model defined in (1.1), and we let $\nu = 1$ and $y_0 = 1$ for computational simplicity. This then corresponds to (2.18) with $\mu = -1/2$, and the corresponding time change is now $\mathbf{Y}_t^{(-1/2)}$. The following theorem gives a representation formula for its density (recall that the function Ψ is given by (2.19)).

Theorem 2.11. *For all $t > 0$, $x > 0$, and $R > 0$,*

$$\varphi_t^{(-\frac{1}{2})}(x) = \frac{(2t)^{-1/4}\,\mathrm{e}^{-t/8}}{2\pi\Gamma(1/4)x^2} \exp\left(\frac{-1}{2tx^2}\right) \int_1^\infty \left\{ \frac{\exp\left(\frac{1}{2tx^2u} - \frac{R}{x\sqrt{u}}\right)}{u^{1/4}(u-1)^{3/4}} \int_{\mathbb{R}} \Psi(t, R + \mathrm{i}s) \exp\left(\frac{\mathrm{i}s}{x\sqrt{u}}\right) \mathrm{d}s \right\} \mathrm{d}u.$$

The theorem follows directly, by the Mellin inversion of the Laplace transform, from Lemma 2.13 below. Before stating the lemma, though, we need to recall the notion of a fractional integral:

Definition 2.12. Let $\alpha \geq 0$ and f be a non-negative Lebesgue measurable function on $(0, \infty)$. Then the fractional integral of order α of the function f is defined, for all $\sigma > 0$, as

$$J_\alpha f(\sigma) := \frac{1}{\Gamma(\alpha)} \int_\sigma^\infty (\tau - \sigma)^{\alpha-1} f(\tau)\mathrm{d}\tau.$$

Lemma 2.13. *For all $\xi > 0$, $t > 0$,*

$$\int_0^\infty u^{-9/4} \varphi_t^{(-1/2)}\left(\frac{1}{u}\right) \mathrm{e}^{-\xi u}\mathrm{d}u = \frac{\mathrm{e}^{-t/8}}{(1+2\xi)^{1/4}} J_{3/4}\Psi(t, \cdot)(\xi).$$

Proof of Lemma 2.13. Dufresne's recurrence formula (see [14] or [23, Section 5.4.5]) allows us to relate the volatility densities for different drifts in the volatility equation in the SABR model; applying [23, Formula (5.92)] with $r = -1/4$, $\mu = -1/2$ and $s = -w < 0$, we obtain

$$\mathrm{e}^{t/8}\int_0^\infty \frac{\lambda^{(-1/2)}(x)}{x^{1/4}}\mathrm{e}^{-wx}\mathrm{d}x = (1+w)^{-1/4}\int_0^\infty \frac{\lambda_t^{(0)}(x)}{x^{1/4}}\mathrm{e}^{-wx}\mathrm{d}x. \tag{2.21}$$

Clearly, $\lambda_t^{(-1/2)}(x) \equiv \frac{1}{2x^2}\varphi_t^{(-1/2)}\left(\frac{1}{2x}\right)$ and $\lambda_t^{(0)}(x) \equiv \frac{1}{2x^2}\varphi_t^{(0)}\left(\frac{1}{2x}\right)$, so that (2.21) reads

$$\mathrm{e}^{t/8}\int_0^\infty \frac{\varphi_t^{(-1/2)}\left(\frac{1}{2x}\right)}{x^{9/4}}\mathrm{e}^{-wx}\mathrm{d}x = (1+w)^{-1/4}\int_0^\infty \frac{\varphi_t^{(0)}\left(\frac{1}{2x}\right)}{x^{9/4}}\mathrm{e}^{-wx}\mathrm{d}x.$$

Substituting $2x = u$ and $\xi = w/2$, we obtain

$$\mathrm{e}^{t/8}\int_0^\infty \frac{\varphi_t^{(-1/2)}(1/u)}{u^{9/4}}\mathrm{e}^{-\xi u}\mathrm{d}u = (1+2\xi)^{-1/4}\int_0^\infty \frac{\varphi_t^{(0)}(1/u)}{u^{9/4}}\mathrm{e}^{-\xi u}\mathrm{d}u. \tag{2.22}$$

We now simplify the integral on the right-hand side of (2.22), using (2.20) and (2.19). Let $\mathcal{L}f(w) \equiv \int_0^\infty f(u)\mathrm{e}^{-wu}\mathrm{d}u$ on $(0,+\infty)$ denote the Laplace transform of the function f. The formula

$$J_\alpha\left(\mathcal{L}f\right)(w) = \int_0^\infty u^{-\alpha}f(u)\mathrm{e}^{-wu}\mathrm{d}u,$$

valid for any $\omega > 0$, is immediate from Definition 2.12, so that (2.20) yields

$$\int_0^\infty \frac{\varphi_t^{(0)}(1/u)}{u^{9/4}}\mathrm{e}^{-\xi u}\mathrm{d}u = \int_0^\infty u^{-3/4}u^{-3/2}\varphi_t^{(0)}\left(\frac{1}{u}\right)\mathrm{e}^{-\xi u}\mathrm{d}u = J_{3/4}\left(\Psi(t,\cdot)\right)(\xi), \tag{2.23}$$

with Ψ defined in (2.19), and the lemma follows from (2.22) and (2.23). $\square$

Remark 2.14. Theorem 2.11 can also be proved in the following way: Dufresne's recurrence formula [23, Theorem 5.25] with $r = -1/4$ and $\beta = 0$ yields

$$\varphi_t^{(-1/2)}\left(\sqrt{y}\right) = \frac{(2t)^{-1/4}\,\mathrm{e}^{-t/8}}{\Gamma(1/4)y}\exp\left(-\frac{1}{2ty}\right)\int_y^\infty \frac{\sqrt{\tau}}{(\tau-y)^{3/4}}\exp\left(\frac{1}{2t\tau}\right)\varphi_t^{(0)}\left(\sqrt{\tau}\right)\mathrm{d}\tau. \tag{2.24}$$

Furthermore, Lemma 2.10 implies

$$\varphi_t^{(0)}\left(\sqrt{\tau}\right) = \frac{\exp\left(-R/\sqrt{\tau}\right)}{2\pi\tau^{3/4}}\int_{-\infty}^\infty \Psi(t,R+\mathrm{i}s)\exp\left(\frac{\mathrm{i}s}{\sqrt{\tau}}\right)\mathrm{d}s,$$

so that, taking this and (2.24) into account, we obtain

$$\begin{aligned}\varphi_t^{(-1/2)}(\sqrt{y}) =& \frac{(2t)^{-1/4}}{2\pi\Gamma(1/4)y}\exp\left(-\frac{t}{8}-\frac{1}{2ty}\right)\\ &\int_y^\infty \frac{\tau^{-1/4}}{(\tau-y)^{3/4}}\exp\left(\frac{1}{2t\tau}-\frac{R}{\sqrt{\tau}}\right)\mathrm{d}\tau\int_{\mathbb{R}}\Psi(t,R+\mathrm{i}s)\exp\left(\frac{\mathrm{i}s}{\sqrt{\tau}}\right)\mathrm{d}s.\end{aligned}$$

The change of variables $\tau = yu$ implies

$$\varphi_t^{(-1/2)}(\sqrt{y}) = \frac{\mathrm{e}^{-t/8}\exp\left(-\frac{1}{2ty}\right)}{2\pi(2t)^{1/4}\Gamma(1/4)y}\int_1^\infty \frac{\exp\left(\frac{1}{2tyu}-\frac{R}{\sqrt{yu}}\right)}{u^{1/4}(u-y)^{3/4}}\mathrm{d}u\int_{\mathbb{R}}\Psi(t,R+\mathrm{i}s)\exp\left(\frac{\mathrm{i}s}{\sqrt{yu}}\right)\mathrm{d}s,$$

and the theorem then follows from the mapping $\sqrt{y}\mapsto x$.

Remark 2.15. If Lemma 2.10 holds also for $R = 0$, then the formula simplifies to

$$\varphi_t^{(-1/2)}(x) = \frac{(2t)^{-1/4}\mathrm{e}^{-t/8}}{2\pi\Gamma(1/4)x^2}\exp\left(-\frac{1}{2tx^2}\right)\int_1^\infty\left(\frac{\exp\left(\frac{1}{2tx^2u}\right)}{u^{1/4}(u-1)^{3/4}}\int_{\mathbb{R}}\Psi(t,\mathrm{i}s)\exp\left(\frac{\mathrm{i}s}{x\sqrt{u}}\right)\mathrm{d}s\right)\mathrm{d}u.$$

2.4.3. *Justification of the Borodin-Salminen formula for $\varphi_t^{(-1/2)}$ via Dufresne's formula.* Using some results of [14], and assuming for simplicity that $y_0 = 1$, we provide a justification for the representation (2.5) for the density $\varphi_t^{(\mu)}$ of $\mathbf{Y}_t^{(\mu)}$, for any $\mu \in \mathbb{R}$:

Proposition 2.16. *The representation* (2.5) *holds.*

Proof. Recall that $\lambda_t^{(\mu)}$ denotes the density of $1/(2\mathbf{Y}_t^{(\mu)})$. Therefore $\varphi_t^{(\mu)}$ clearly satisfies

$$\varphi_t^{(\mu)}(x) = \frac{1}{2x^2}\lambda_t^{(\mu)}\left(\frac{1}{2x}\right), \qquad \text{for all } x > 0. \tag{2.25}$$

Dufresne [14, Theorem 4.2] showed that $\lambda_t^{(\mu)}$ admits the following closed-form representation:

$$\lambda_t^{(\mu)}(x) = \exp\left(-\frac{\mu^2 t}{2}\right) p_t^{(\mu)}(x), \tag{2.26}$$

where

$$p_t^{(\mu)}(x) \equiv 2^{-\mu} x^{-\frac{\mu+1}{2}} \int_{\mathbb{R}} \mathrm{e}^{-x\cosh^2(y)} q(y,t) \cos\left[\frac{\pi}{2}\left(\frac{y}{t} - \mu\right)\right] \mathrm{H}_\mu\left(\sqrt{x}\sinh(y)\right) \mathrm{d}y,$$

H_μ is the Hermite function of order μ, and $q(y,t) \equiv \frac{1}{\pi\sqrt{2t}} \exp\left(\frac{\pi^2}{8t} - \frac{y^2}{2t}\right)\cosh(y)$. When $\mu \neq -2, -4, \ldots$, Dufresne [14, Equation (4.13)] proved the following equivalent formulation:

$$p_t^{(\mu)}(x) = \frac{2\Gamma(1+\mu/2)}{\Gamma(3/2)x^{-\mu/2}} \int_0^\infty \mathrm{e}^{-x\cosh^2(y)} q(y,t)\sinh(y)\sin\left(\frac{\pi y}{2t}\right) \mathrm{M}\left(\frac{1-\mu}{2}, \frac{3}{2}; x\sinh^2(y)\right) \mathrm{d}y, \tag{2.27}$$

where M is the confluent hypergeometric function (Kummer function), also denoted by ${}_1F_1$. We shall call this formula 'Dufresne's formula'. Since $\Gamma(3/2) = \sqrt{\pi}/2$, the formulae (2.25), (2.26)), and (2.27)) with $\mu = -1/2$, yield, after simplifications,

$$\begin{aligned} \varphi_t^{(-1/2)}(x) = & \frac{2^{1/4}\Gamma(3/4)}{\pi^{3/2}t^{1/2}} x^{-9/4} \exp\left(\frac{\pi^2}{8t} - \frac{t}{8} - \frac{1}{2x}\right) \\ & \int_0^\infty \exp\left(-\frac{y^2}{2t}\right)\sinh(y)\cosh(y)\sin\left(\frac{\pi y}{2t}\right) \mathrm{M}\left(\frac{3}{4}, \frac{3}{2}; -\frac{\sinh^2(y)}{2x}\right)\mathrm{d}y. \end{aligned} \tag{2.28}$$

From [1, Formula 13.1.27], the identity $M(a,b;z) = \mathrm{e}^z M(b-a, b, -z)$ holds, and hence

$$\begin{aligned} \mathrm{M}\left(\frac{3}{4}, \frac{3}{2}; -\frac{\sinh^2(y)}{2x}\right) &= \exp\left(-\frac{\sinh^2(y)}{2x}\right)\mathrm{M}\left(\frac{3}{4}, \frac{3}{2}; \frac{\sinh^2(y)}{2x}\right) \\ &= \mathrm{e}^{1/(4x)} \exp\left(-\frac{\cosh(2y)}{4x}\right) M\left(\frac{3}{4}, \frac{3}{2}; \frac{\sinh^2(y)}{2x}\right), \end{aligned}$$

so that (2.28) simplifies to

$$\begin{aligned} \varphi_t^{(-1/2)}(x) = & \frac{2^{1/4}\Gamma(3/4)}{\pi^{3/2}\sqrt{t}} \exp\left(\frac{\pi^2}{8t} - \frac{t}{8} - \frac{1}{4x}\right) x^{-9/4} \\ & \int_0^\infty \exp\left(-\frac{\cosh(2y)}{4x}\right)\exp\left(-\frac{y^2}{2t}\right)\sinh(y)\cosh(y)\sin\left(\frac{\pi y}{2t}\right)\mathrm{M}\left(\frac{3}{4}, \frac{3}{2}; \frac{\sinh^2(y)}{2x}\right)\mathrm{d}y, \end{aligned}$$

which yields (2.5)-(2.6) with $\mu = -3/4$. □

2.4.4. *Alternative representation of the SABR-density and Yor's formula.* We now provide an alternative representation for the quantity $\mathbb{P}\left(\int_0^t \mathrm{e}^{2\nu Z_s^{(-\nu/2)}}\mathrm{d}s \in \mathrm{d}\widetilde{r}\right)$ appearing in (2.5) in terms of the modified Bessel function of the second kind K (see [1, Section 9.6]). This representation is derived from the results obtained by Yor in [47] (see also [36]), and we use here the formulation given in [23, Section 4.7].

Proposition 2.17. *For all $t > 0$ and $y > 0$,*

$$\varphi_t^{(-1/2)}(y) = \frac{y_0^2}{2\pi^{3/2}\nu^3} t^{-\frac{1}{2}} y^{-2} \exp\left(\frac{\pi^2}{2\nu^2 t} - \frac{\nu^2 t}{8} - \frac{y_0^2}{2\nu^2 y}\right)$$
$$\times \int_0^\infty \sinh(u)\sqrt{\cosh(u)} \mathrm{e}^{\frac{y_0^2}{4\nu^2 y}\cosh(u)^2} \mathrm{K}_{1/4}\left(\frac{y_0^2}{4\nu^2 y}\cosh(u)^2\right) \exp\left(-\frac{u^2}{2\nu^2 t}\right) \sin\left(\frac{\pi u}{\nu^2 t}\right) \mathrm{d}u.$$

Proof. An explicit formula for the joint density μ_t of the random variables

$$\mathbf{Y}_t^{(-1/2)} = \int_0^t Y_s^2 \mathrm{d}s \qquad \text{and} \qquad Y_t = y_0 \exp\left(\nu Z_t^{(-\nu/2)}\right)$$

can be found in [23, Formula (4.84)]. It follows from this formula that

$$\mu_t(y,z) = \frac{y_0^{3/2}\exp\left(-\frac{\nu^2 t}{8} + \frac{\pi^2}{2\nu^2 t} - \frac{y_0^2+z^2}{2\nu^2 y}\right)}{\sqrt{2zt}\pi^{3/2}\nu^3 y^2} \int_0^\infty \exp\left(-\frac{u^2}{2\nu^2 t} - \frac{y_0 z}{\nu^2 y}\cosh(u)\right)\sinh(u)\sin\left(\frac{\pi u}{\nu^2 t}\right)\mathrm{d}u.$$

Integrating out the variable z, we get the following expression for the density $\varphi_t^{(-1/2)}$ of $\mathbf{Y}_t^{(-1/2)}$:

$$\varphi_t^{(-1/2)}(y) = \frac{y_0^{3/2}\exp\left(-\frac{\nu^2 t}{8} + \frac{\pi^2}{2\nu^2 t} - \frac{y_0^2}{2\nu^2 y}\right)}{\sqrt{2t}\pi^{\frac{3}{2}}\nu^3 y^2} \int_0^\infty \exp\left(-\frac{u^2}{2\nu^2 t}\right)\sin\left(\frac{\pi u}{\nu^2 t}\right)\sinh(u)\mathrm{d}u$$
$$\times \int_0^\infty \exp\left(-\frac{z^2}{2\nu^2 y} - \frac{y_0 z}{\nu^2 y}\cosh(u)\right)\frac{\mathrm{d}z}{\sqrt{z}}. \tag{2.29}$$

We call (2.29) 'Yor's formula', and we now simplify the last integral, which we denote J. Clearly,

$$J = 2\int_0^\infty \exp\left(-\frac{1}{2\nu^2 y}x^4 - \frac{y_0\cosh(u)}{\nu^2 y}x^2\right)\mathrm{d}x.$$

Applying the formula [20, 3.323 (3)])

$$\int_0^\infty \exp\left(-\beta^2 x^4 - 2\gamma^2 x^2\right)\mathrm{d}x = 2^{-3/2}\frac{\gamma}{\beta}\exp\left(\frac{\gamma^4}{2\beta^2}\right)\mathrm{K}_{1/4}\left(\frac{\gamma^4}{2\beta^2}\right),$$

which holds for all β and γ with $|\arg\beta| < \frac{\pi}{4}$ and $|\arg\gamma| < \frac{\pi}{4}$, with $\beta = 1/(\nu\sqrt{2y})$ and $\gamma = \sqrt{y_0\cosh(u)}/(\nu\sqrt{2y})$, we obtain

$$J = \sqrt{\frac{y_0\cosh(u)}{2}}\exp\left(\frac{y_0^2}{4\nu^2 y}\cosh(u)^2\right)\mathrm{K}_{1/4}\left(\frac{y_0^2}{4\nu^2 y}\cosh(u)^2\right),$$

and the proposition follows from this and (2.29). □

It is also possible to represent the density $\varphi_t^{(-1/2)}$ of $\mathbf{Y}_t^{(-1/2)}$ by an integral formula using the Kummer function of the second kind U [1, Section 13, Equation 13.1.3]:

Lemma 2.18. *For all $t > 0$ and $y > 0$,*

$$\varphi_t^{(-1/2)}(y) = \frac{y_0^{5/2}}{2^{5/4}\pi\nu^{7/2}} t^{-1/2} y^{-9/4}\exp\left(\frac{\pi^2}{2\nu^2 t} - \frac{\nu^2 t}{8} - \frac{y_0^2}{2\nu^2 y}\right)$$
$$\times \int_0^\infty \mathrm{U}\left(\frac{3}{4},\frac{3}{2},\frac{y_0^2}{2\nu^2 y}\cosh(u)^2\right)\sinh(u)\cosh(u)\exp\left(-\frac{u^2}{2\nu^2 t}\right)\sin\left(\frac{\pi u}{\nu^2 t}\right)\mathrm{d}u.$$

Proof. Recall that [1, Equation 13.6.21]

$$\mathrm{U}(a,2a,x) = \frac{1}{\sqrt{\pi}}\exp\left(\frac{x}{2}\right)x^{\frac{1}{2}-a}\mathrm{K}_{a-\frac{1}{2}}\left(\frac{x}{2}\right),$$

for all $a > 0$ $x > 0$. Therefore,

$$\mathrm{K}_{1/4}\left(\frac{y_0^2}{4\nu^2 y}\cosh(u)^2\right) = \sqrt{\pi}\exp\left(-\frac{y_0^2}{4\nu^2 y}\cosh(u)^2\right)\left[\frac{y_0^2}{2\nu^2 y}\cosh(u)^2\right]^{1/4}\mathrm{U}\left(\frac{3}{4},\frac{3}{2},\frac{y_0^2}{2\nu^2 y}\cosh(u)^2\right),$$

and the lemma follows from Proposition 2.17. □

Remark 2.19. Dufresne's formula (2.27) and Yor's formula (2.29) yield different integral representations for the density $\varphi_t^{(-1/2)}$. It is not obvious, however, to show directly that they are the same, and we refer the reader to some hints in [36, Section 4.3].

3. Mass at zero for the correlated SABR model

The present section deals with the behaviour of the mass at zero in the correlated ($\rho \neq 0$) SABR model, and in the associated model for the Brownian motion in the SABR plane. We shall consider the following modified version of the SABR model:

$$\begin{aligned} \mathrm{d}X_t &= Y_t X_t^\beta \mathrm{d}W_t + \frac{\beta}{2} Y_t^2 X_t^{2\beta-1}\mathrm{d}t, && X_0 = x_0 > 0, \\ \mathrm{d}Y_t &= \nu Y_t \mathrm{d}Z_t, && Y_0 = y_0 > 0, \\ \mathrm{d}\langle Z, W\rangle_t &= \rho \mathrm{d}t, && \end{aligned} \tag{3.1}$$

with $\nu > 0$, $\rho \in (-1,1)$, $\beta \in [0,1)$. Note that the behaviour of the drift in the first stochastic differential equation and its implications for the mass at zero of the modified model are significantly different in the cases $0 < \beta < 1/2$ and $1/2 < \beta < 1$: in the former case the drift explodes when the process X approaches zero, while it vanishes in the latter case. In the case $\beta = 1/2$, the drift does not depend on the process X. The model in (3.1) is a modification of the SABR model which characterises a Brownian motion in a suitably chosen Riemannian manifold with boundary (the SABR plane [27, Subsection 3.2]), see Lemma 3.3 below. Of particular interest are two special cases of (3.1): the uncorrelated case $\rho = 0$ and the $\beta = 0$ case. Computing the mass at zero for (3.1) in the uncorrelated case sheds light on the influence of the drift, by a comparison with the results obtained in Section 2. In the case $\beta = 0$, the drift in (3.1) vanishes, and the model coincides with the original SABR model (1.1) with $\beta = 0$. In fact, according to [6] and [27], in the prevalence of low interest rates, the choice $\beta = 0$ is essential. Note that up to a deterministic time change, we can assume $\nu = 1$, which we shall tacitly do without loss of generality from now on if not stated otherwise.

3.1. **SABR geometry and geometry preserving mappings.** We first exhibit a set of mappings allowing to translate the properties of one SABR model to another. Let $\mathbb{H} = \{x + \mathrm{i}y : x \in \mathbb{R}, y > 0\}$ and $\mathbb{H}_+ := \{x+\mathrm{i}y : (x,y) \in (0,\infty)^2\}$. $(\mathbb{H}, g)$ will denote the classical Poincaré plane with its associated Riemannian metric [21, Section 3.9], and $(\mathcal{S}, g)$ the general SABR plane (generated by (3.1)). As mentioned above, of particular interest are the two cases of the uncorrelated SABR plane, denoted by $(\mathbb{U}, u)$, and of $(\mathcal{S}^0, g^0)$, the general SABR plane with $\beta = 0$. Note that only $\mathbb{U}$ and $\mathcal{S}$ exhibit a drift. We also denote by $\mathcal{S}^0_+ := \mathcal{S}^0 \cap \{x + \mathrm{i}y : (x,y) \in (0,\infty)^2\}$. The following tensors, h, g, g^0 and u generate Riemannian metrics on their respective spaces:

$$h(\widetilde{x},\widetilde{y}) = \frac{\mathrm{d}\widetilde{x}^2}{\widetilde{y}^2} + \frac{\mathrm{d}\widetilde{y}^2}{\widetilde{y}^2}, \quad (\widetilde{x},\widetilde{y}) \in \mathbb{R}\times(0,\infty), \tag{3.2}$$

$$g(x,y) = \frac{\mathrm{d}x^2}{\bar{\rho}^2 y^2 x^{2\beta}} - \frac{2\rho \mathrm{d}x\mathrm{d}y}{\bar{\rho}^2 y^2 x^\beta} + \frac{\mathrm{d}y^2}{\bar{\rho}^2 y^2}, \quad (x,y) \in (0,\infty)^2, \tag{3.3}$$

$$g^0(\widehat{x},\widehat{y}) = \frac{\mathrm{d}\widehat{x}^2}{\bar{\rho}^2\widehat{y}^2} - \frac{2\rho\mathrm{d}\widehat{x}\mathrm{d}\widehat{y}}{\bar{\rho}^2\widehat{y}^2} + \frac{\mathrm{d}\widehat{y}^2}{\bar{\rho}^2\widehat{y}^2}, \quad (\widehat{x},\widehat{y}) \in \mathbb{R}\times(0,\infty), \tag{3.4}$$

$$u(\bar{x},\bar{y}) = \frac{\mathrm{d}\bar{x}^2}{\bar{y}^2\bar{x}^{2\beta}} + \frac{\mathrm{d}\bar{y}^2}{\bar{y}^2} \quad (\bar{x},\bar{y}) \in (0,\infty)^2, \tag{3.5}$$

and the following diagram summarises the different relations between the mappings and the spaces (we also include the corresponding coordinate notations):

$$
\begin{tikzcd}
 & \overset{(\widetilde{x},\widetilde{y})}{\mathbb{H}} \arrow[dr, "\bar{\chi}"] & \\
\overset{(\widehat{x},\widehat{y})}{\mathcal{S}^0} \arrow[ur, bend left, "\widetilde{\phi}_0"] \arrow[dr, "\chi"] & & \overset{(\bar{x},\bar{y})}{\mathbb{U}} \arrow[ul, bend right, "\widetilde{\varphi}^0"'] \\
 & \overset{(x,y)}{\mathcal{S}} \arrow[uu, "\widetilde{\phi}_0^0"] \arrow[ul, bend left, "\widehat{\varphi}^0"] \arrow[ur, bend right, "\bar{\phi}_0"'] &
\end{tikzcd}
$$

Regarding the mapping notations, subscripts are related to the correlation parameter (ϕ_0 'annihilates' ρ), whereas superscripts ^{0} indicate that the map 'annuls' the parameter β; the map χ reintroduces this parameter. The mappings between these spaces are defined as follows:

$$
\begin{array}{rrcl}
\widetilde{\phi}_0^0: & \mathcal{S} & \longrightarrow & \mathbb{H}, \\
 & (x,y) & \longmapsto & (\widetilde{x},\widetilde{y}) := \left(\dfrac{x^{1-\beta}}{\overline{\rho}(1-\beta)} - \dfrac{\rho y}{\overline{\rho}}, y\right), \\
\widehat{\varphi}^0: & \mathcal{S} & \longrightarrow & \mathcal{S}^0, \\
 & (x,y) & \longmapsto & (\widehat{x},\widehat{y}) := \left(\dfrac{x^{1-\beta}}{1-\beta}, y\right), \\
\bar{\phi}_0: & \mathcal{S} & \longrightarrow & \mathbb{U}, \\
 & (x,y) & \longmapsto & (\bar{x},\bar{y}) = \left((1-\beta)^{\frac{1}{1-\beta}}\left(\dfrac{x^{1-\beta}}{\overline{\rho}(1-\beta)} - \dfrac{\rho y}{\overline{\rho}}\right)^{\frac{1}{1-\beta}}, y\right), \ \rho \le 0, \\
\widetilde{\phi}_0: & \mathcal{S}^0_+ & \longrightarrow & \mathbb{H}, \\
 & (\widehat{x},\widehat{y}) & \longmapsto & (\widetilde{x},\widetilde{y}) := \left(\dfrac{\widehat{x}-\rho\widehat{y}}{\overline{\rho}}, \widehat{y}\right), \\
\chi: & \mathcal{S}^0_+ & \longrightarrow & \mathcal{S}, \\
 & (\widehat{x},\widehat{y}) & \longmapsto & (x,y) := \left((1-\beta)^{\frac{1}{1-\beta}}\widehat{x}^{\frac{1}{1-\beta}}, \widehat{y}\right), \\
\bar{\chi}: & \mathbb{H}_+ & \longrightarrow & \mathbb{U}, \\
 & (\widetilde{x},\widetilde{y}) & \longmapsto & (\bar{x},\bar{y}) := \left((1-\beta)^{\frac{1}{1-\beta}}\widetilde{x}^{\frac{1}{1-\beta}}, \widetilde{y}\right), \\
\widetilde{\varphi}^0: & \mathbb{U} & \longrightarrow & \mathbb{H}_+, \\
 & (\bar{x},\bar{y}) & \longmapsto & (\widetilde{x},\widetilde{y}) := \left(\dfrac{\bar{x}^{1-\beta}}{1-\beta}, \bar{y}\right).
\end{array}
\tag{3.6}
$$

From now on, if not indicated otherwise, we restrict the domains of the above maps to the first quadrant $(0,\infty)^2$, which–when considering compositions–impose restrictions on the parameters in order to ensure that images also belong to this set (for example the restriction $\rho \in (-1,0]$ needs to be imposed for the composition $\phi_0^0 \circ \chi$). While the map ϕ_0 can be extended to the whole upper halfplane $\mathbb{R}\times(0,\infty)$, thus describing an asset with negative value, the maps φ^0, ϕ_0^0 and χ are in general not meaningful there. They can be extended to $x = 0$ though, and are non-differentiable there. The following theorem gathers the properties of all these maps:

Theorem 3.1. *The diagram is commutative and all the mappings in* (3.6) *are local isometries on their respective spaces:*

- *the maps $\widehat{\varphi}^0$ and χ (resp. $\widetilde{\varphi}^0$ and $\bar{\chi}$) on $(0,\infty)^2$ are onto and inverse to one another;*
- *the compositions $\bar{\phi}_0 \circ \widetilde{\varphi}^0$ and $\widehat{\varphi}^0 \circ \widetilde{\phi}_0$ coincide with $\widetilde{\phi}_0^0$;*
- *it holds that $\chi \circ \widetilde{\phi}_0^0 = \widetilde{\phi}_0$ and $\widetilde{\phi}_0^0 \circ \bar{\chi} = \bar{\phi}_0$ and the latter is well defined for $\rho \in (-1,0]$;*
- *the map $\widehat{\varphi}^0$ (resp. $\widetilde{\varphi}^0$) transforms Brownian motion on $(\mathcal{S},g)$ (resp. $(\mathbb{U},u)$) into the SABR model* (1.1) *with $\beta = 0$ (resp. $\rho,\beta = 0$), which in turn is transformed back to Brownian motion on its original spaces by the map χ (resp. $\bar{\chi}$);*
- *the maps $\bar{\phi}_0$ (resp. the extension of $\widetilde{\phi}_0$) transforms Brownian motion on $(\mathcal{S},g)$ (resp. $(\mathcal{S}^0,g^0)$) into its uncorrelated version on $(\mathbb{U},u)$ (resp. $(\mathbb{H},h)$).*

Remark 3.2. The map $\widetilde{\phi}_0^0$ was first considered in [27], where it was observed there that it is a local isometry mapping Brownian motion on $(\mathcal{S}, g)$ to a Brownian motion on the hyperbolic half-plane $(\mathbb{H}, h)$.

Proof. The first three items follows from simple computation; the remaining statements follow from Lemmas (3.3),(3.5), (3.6) and (3.7) below. $\square$

Lemma 3.3. *The process (X, Y) with dynamics (3.1) coincides in law with Brownian motion on the manifold $(\mathcal{S}, g)$. We define the process $(\widehat{X}, \widehat{Y})$ pathwise by applying to (X, Y) the space transformation $\widehat{\varphi}^0 : \mathcal{S} \to \mathcal{S}^0$, i.e. by setting*

$$(\widehat{X}_t, \widehat{Y}_t) := \left(\frac{X_t^{1-\beta}}{1-\beta}, Y_t\right), \quad \text{for all } t \geq 0. \tag{3.7}$$

Then $(\widehat{X}, \widehat{Y})$ is a SABR process with $\beta = 0$. Furthermore, the process (3.7) coincides in law with Brownian motion on the manifold $(\mathcal{S}^0, g^0)$, to which we refer as the correlated hyperbolic plane.

Proof. The statement that (3.1) has the same law as Brownian motion on $(\mathcal{S}, g)$ is verified by computing the infinitesimal generator of (3.1), which coincides with the Laplace-Beltrami operator $\frac{1}{2}\Delta_g$ on a manifold with metric tensor $g(x, y)$ (see (B.5) and (B.6) for more detail). The second statement is straightforward from Itô's formula, which transforms the system (3.1) into

$$\begin{aligned} \mathrm{d}\widehat{X}_t &:= \widehat{Y}_t \mathrm{d}W_t, & \widehat{X}_0 &= \widehat{x}_0 := x_0^{1-\beta}/(1-\beta), \\ \mathrm{d}\widehat{Y}_t &= \nu \widehat{Y}_t \mathrm{d}Z_t, & \widehat{Y}_0 &= \widehat{y}_0, \\ \mathrm{d}\langle W, Z\rangle_t &= \rho \mathrm{d}t. \end{aligned} \tag{3.8}$$

It is easy to see that (3.8) has SABR dynamics (1.1) with parameters $\beta = 0$, and $\rho \in (-1, 1)$. The generator of (3.8) coincides with the Laplace-Beltrami operator $\frac{1}{2}\Delta_{g^0}$ of the respective manifold, which yields the last statement. $\square$

Remark 3.4. Since $\beta \in [0, 1)$, we have

$$\mathbb{P}_\infty = \mathbb{P}(X_t = 0 \text{ for some } t \in (0, \infty)) = \mathbb{P}(\widehat{X}_t = 0 \text{ for some } t \in (0, \infty)). \tag{3.9}$$

Note that the map $\widehat{\varphi}^0 : \mathcal{S} \to \mathcal{S}^0$ is applied in the proof of Theorem 3.9 below.

Lemma 3.5. *$\widetilde{\phi}_0 : \mathcal{S}^0 \to \mathbb{H}$ is a global isometry and transforms the SABR model (1.1) with $\beta = 0$ into a Brownian motion on $(\mathbb{H}, h)$. Furthermore, the heat (or transition) kernel of the solution of the system (3.8) is available in closed form:*

$$\frac{1}{\overline{\rho}} K^h_{\phi_0(x,y)}(s, \phi_0(x, y)), \qquad \text{for } s > 0, (x, y) \in \mathcal{S}^0,$$

where $K^h_{(\widetilde{x},\widetilde{y})}(s, \cdot)$ denotes the hyperbolic heat kernel at $(\widetilde{x}, \widetilde{y}) \in \mathbb{H}$, for which a closed-form expression as well as short- and large-time asymptotics are known (see [21, Equation (9.35)] *and* [27, Appendix]*).*

Proof. The following shows that ϕ_0 is in fact a global isometry: $\widetilde{\phi}_0$ is onto and invertible on $\mathcal{S}^0$ and, for any $(x, y) \in \mathcal{S}^0$, its Jacobian

$$\nabla \widetilde{\phi}_0(x, y) = \begin{pmatrix} 1/\overline{\rho} & -\rho/\overline{\rho} \\ 0 & 1 \end{pmatrix},$$

is independent of x and does not explode at $x = 0$. Furthermore, for any $(x, y) \in \mathcal{S}^0$,

$$\left(\widetilde{\phi}_0^* h\right)(x, y) = \widetilde{\phi}_0^* \left(\frac{\mathrm{d}\widetilde{x}^2 + \mathrm{d}\widetilde{y}^2}{\widetilde{y}^2}\right) = \frac{1}{y^2}\left(\frac{\mathrm{d}x}{\overline{\rho}} - \frac{\rho \mathrm{d}y}{\overline{\rho}}\right)^2 + \frac{(\mathrm{d}y)^2}{y^2} = g^0(x, y).$$

The last statement follows from Lemma B.5 together with $\det(\nabla \widetilde{\phi}_0(\cdot)) = 1/\overline{\rho} \neq 0$. One can easily verify by Itô's lemma that the dynamics (3.8) for general $\rho \in (-1, 1)$ are transformed into (3.8) for $\rho = 0$ under the map $\widetilde{\phi}_0$. $\square$

Lemma 3.6. *The map χ (resp. $\bar{\chi}$) is a local isometry between $(\mathcal{S}^0_+, g^0)$ and $(\mathcal{S}, g)$ (resp. $(\mathbb{H}_+, h)$ and $(\mathbb{U}, u)$) and transforms the Brownian motion on the hyperbolic plane $(\mathcal{S}^0_+, g^0)$ (resp. $(\mathbb{H}_+, h)$), whose dynamics are described by* (3.8)*, into a Brownian motion on the general SABR plane $(\mathcal{S}, g)$ (resp. $(\mathbb{U}, u)$), satisfying* (3.1).

Proof. For a local isometry between $(\mathcal{S}^0_+, g^0)$ and $(\mathcal{S}, g)$ (resp. $(\mathbb{H}, h)$ and $(\mathbb{U}, u)$), it holds that for any $(\widehat{x}, \widehat{y}) \in \mathcal{S}^0$ (resp.$(\widetilde{x}, \widetilde{y}) \in \mathbb{H}$) there exists a small open neighbourhood $U_{(\widehat{x},\widehat{y})} \subset \mathcal{S}^0$ (resp. $U_{(\widetilde{x},\widetilde{y})} \subset \mathbb{H}$), such that the map $\chi_{|U_{(\widehat{x},\widehat{y})}}$ (resp. $\bar{\chi}_{|U_{(\widetilde{x},\widetilde{y})}}$) is an isometry onto its image, in particular it satisfies the pullback relation

$$(\chi^* g)(x, y) = \chi^* \left(\frac{\mathrm{d}x^2}{\overline{\rho} x^{2\beta} y^2} + \frac{2\rho \mathrm{d}x \mathrm{d}y}{\overline{\rho} x^{\beta} y^2} + \frac{\mathrm{d}y^2}{\overline{\rho} y^2} \right) = \frac{\mathrm{d}\widehat{x}^2 + 2\rho \mathrm{d}\widehat{x}\mathrm{d}\widehat{y} + \mathrm{d}\widehat{y}^2}{\overline{\rho}\widehat{y}^2} = g^0(\widehat{x}, \widehat{y}),$$

respectively, for zero correlation

$$(\bar{\chi}^* u)(\bar{x}, \bar{y}) = \bar{\chi}^* \left(\frac{\mathrm{d}\bar{x}^2}{\bar{x}^{2\beta} \bar{y}^2} + \frac{\mathrm{d}\bar{y}^2}{\bar{y}^2} \right) = \frac{\mathrm{d}\widetilde{x}^2 + \mathrm{d}\widetilde{y}^2}{\widetilde{y}^2} = h(\widetilde{x}, \widetilde{y}).$$

For any $(\widehat{x}, \widehat{y}) \in \mathcal{S}^0$ (resp. $(\widetilde{x}, \widetilde{y}) \in \mathbb{H}$), the Jacobians read

$$\nabla\chi(\widehat{x}, \widehat{y}) = \begin{pmatrix} (1-\beta)^{\frac{\beta}{1-\beta}} \widehat{x}^{\frac{\beta}{1-\beta}} & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \nabla\bar{\chi}(\widetilde{x}, \widetilde{y}) = \begin{pmatrix} (1-\beta)^{\frac{\beta}{1-\beta}} \widetilde{x}^{\frac{\beta}{1-\beta}} & 0 \\ 0 & 1 \end{pmatrix},$$

respectively, hence the local pullback property is clearly satisfied by χ (resp. $\bar{\chi}$). The last statement follows by Itô's lemma. ∎

We now verify that $\bar{\phi}_0$ is a 'geometry-preserving' map from the general SABR plane $(\mathcal{S}, g)$ into the uncorrelated SABR plane $(\mathbb{U}, u)$, which of course reduces to the identity map when $\rho = 0$, and to $\widetilde{\phi}_0$ when $\beta = 0$.

Lemma 3.7. *For any $\rho \in (-1, 0]$ and any $(x, y) \in \mathcal{S}$, the space transformation $\bar{\phi}_0 : \mathcal{S} \longrightarrow \mathbb{U}$ in* (3.6) *is a local isometry between $(\mathcal{S}, g)$ and $(\mathbb{U}, u)$.*

Proof. The statement follows directly from the fact that the map $\bar{\phi}_0$ and its partial derivatives

$$\begin{aligned}
\partial_x \bar{x}(x, y) &= \frac{x^{-\beta}}{\overline{\rho}} (1-\beta)^{\frac{\beta}{1-\beta}} \left(\frac{x^{1-\beta}}{\overline{\rho}(1-\beta)} - \frac{\rho y}{\overline{\rho}} \right)^{\beta/(1-\beta)}, \\
\partial_y \bar{x}(x, y) &= -\frac{\rho}{\overline{\rho}} (1-\beta)^{\frac{\beta}{1-\beta}} \left(\frac{x^{1-\beta}}{\overline{\rho}(1-\beta)} - \frac{\rho y}{\overline{\rho}} C \right)^{\beta/(1-\beta)}, \\
\partial_x \bar{y}(x, y) &= 0, \qquad \partial_y \bar{y}(x, y) = 1,
\end{aligned} \tag{3.10}$$

satisfy the following system of differential equations implied by the local pullback property $\left(\bar{\phi}_0^* u\right)(\bar{x}, \bar{y}) = g(x, y)$, for any $(x, y) \in \mathcal{S}$, $(\bar{x}, \bar{y}) \in \mathbb{U}$ for the Riemannian metrics g and u:

$$\begin{cases}
\dfrac{(\partial_x \bar{x})^2}{\bar{x}^{2\beta} \bar{y}^2} + \dfrac{(\partial_x \bar{y})^2}{\bar{y}^2} = \dfrac{1}{\overline{\rho}^2 y^2 x^{2\beta}}, \\
\dfrac{2(\partial_x \bar{x} \partial_y \bar{x})}{\bar{x}^{2\beta} \bar{y}^2} + \dfrac{2(\partial_x \bar{y} \partial_y \bar{y})}{\bar{y}^2} = \dfrac{-2\rho}{\overline{\rho}^2 y^2 x^{\beta}}, \\
\dfrac{(\partial_y \bar{x})^2}{\bar{x}^{2\beta} \bar{y}^2} + \dfrac{(\partial_y \bar{y})^2}{\bar{y}^2} = \dfrac{1}{\overline{\rho}^2 y^2}.
\end{cases}$$

∎

As an application of Lemma 3.7 it may be possible to relate the absolutely continuous part of the distribution of Brownian motion on the uncorrelated SABR plane $(\mathbb{U}, u)$ and that of Brownian motion on the general SABR plane $(\mathcal{S}, g)$ via the relation (B.4) of the heat kernels [44]; this can be performed following similar steps as in [27], but care is needed, as discussed below.

Lemma 3.8. *Let K^g and K^u denote the fundamental solutions (in terms of Lebesgue) of the heat equations corresponding to the metrics g and u, then, for any $z = (x, y) \in \mathcal{S}$,*

$$K_Z^g(s,z) = \frac{(1-\beta)^{\frac{\beta}{1-\beta}}}{\overline{\rho}x^\beta}\left(\frac{x^{1-\beta}}{\overline{\rho}(1-\beta)} - \frac{\rho y}{\overline{\rho}}\right)^{\frac{\beta}{1-\beta}} K^u_{\bar{\phi}_0(z)}(s, \phi_0^0(z)). \tag{3.11}$$

When $\beta = 1/2$, the formulae simplify to $\bar{\phi}_0(x,y) \equiv \left(\frac{1}{(1-\rho)^2}\left(x - \sqrt{x}\rho y + \frac{\rho^2 y^2}{4}\right), y\right)$, and $\det(\nabla\bar{\phi}_0(x,y)) = \left(1 - \frac{\rho y}{2\sqrt{x}}\right)/(1-\rho)^2$, for all $(x,y) \in \mathcal{S}$.

Proof. The statement follows from Lemma B.5: the Radon-Nikodym derivatives are $\frac{\mathrm{d}z}{\mathrm{d}\mu_g(z)} = \overline{\rho}^2 y^2 x^\beta$ and $\frac{\mathrm{d}\bar{z}}{\mathrm{d}\mu_u(\bar{z})} = \bar{y}^2\bar{x}^\beta$, with μ_g and μ_u the Riemannian volume elements on $\mathcal{S}$ and $\mathbb{U}$ (Definition B.2 in Appendix B), and the Jacobian of $\bar{\phi}_0$ at $z = (x,y) \in \mathcal{S}$ is as in (3.10), so that

$$\det\left(\nabla\bar{\phi}_0(x,y)\right) = \frac{(1-\beta)^{\frac{\beta}{1-\beta}}}{\overline{\rho}x^\beta}\left(\frac{x^{1-\beta}}{\overline{\rho}(1-\beta)} - \frac{\rho y}{\overline{\rho}}\right)^{\frac{\beta}{1-\beta}}.$$

□

Such a relation of heat kernels relies on the property B.1 of Laplace-Beltrami operators, which is not meaningful for (B.5) at $x = 0$ for general β. Hence a statement relating the heat kernels might not hold true in the vicinity of the origin. Although in the case of exploding Jacobians the relation (B.4) of 'kernels' formally indicates that the map under consideration induces an atom, it does not allow for an exact computation. Remark further, that non-differentiability issues at $x = 0$ of the maps may induce a local time at this point, which we do not investigate further for $\widehat{\varphi}^0$, $\widetilde{\varphi}^0$, $\bar{\chi}$ and χ, as we imposed Dirichlet boundary conditions at $x = 0$. But they might be of importance for the map $\bar{\phi}_0$ introduced in (3.6) above and for the map $\widetilde{\phi}_0^0$ considered in [27]. A statement similar to Lemma 3.8 below was made in [27] relating K^g to the hyperbolic heat kernel K^h; in their analysis, the determinant was $\det(\nabla\widetilde{\phi}_0^0(x,y)) \equiv x^{-\beta}/\rho$.

3.2. **Application: Large-time behaviour of the mass.** We now compute the large-time limit of the mass at zero in the modified SABR model (3.1), with correlation:

$$\mathbb{P}_\infty =: \lim_{t\uparrow\infty}\mathbb{P}(X_t = 0). \tag{3.12}$$

The computation of the mass (Theorem 3.9 below) follows the works of Hobson [28] on time changes. We apply such a technique to progress from the Brownian motion on the correlated hyperbolic plane (3.8) to a correlated Brownian motion on the Euclidean plane. The joint distribution of hitting times of zero of two (correlated) Brownian motions without drift was first established by Iyengar [30], and refined by Metzler [38] (see also [9] for further results on hitting times of correlated Brownian motions). We also borrow some ideas from [13], where Hobson's construction for the normal SABR model [28, Example 5.2] is extended to (3.1) for general $\beta \in [0,1]$. This indeed follows from the observation that stochastic time change methods, going back to Volkonskii [46], can still be applied to the Brownian motion on the SABR plane. In order to formulate our next statement, we introduce several auxiliary parameters (see [38]):

$$a_1 := \frac{x_0^{1-\beta}}{1-\beta}, \qquad a_2 := \frac{y_0}{\nu}, \qquad r_0 := \sqrt{\frac{a_1^2 + a_2^2 - 2\rho a_1 a_2}{\overline{\rho}^2}},$$

$$\alpha := \begin{cases} \pi + \arctan(-\overline{\rho}/\rho), & \text{if } \rho > 0,\\ \frac{\pi}{2}, & \text{if } \rho = 0,\\ \arctan(-\overline{\rho}/\rho), & \text{if } \rho < 0,\end{cases} \qquad \theta_0 := \begin{cases} \pi + \arctan(a_2\overline{\rho}/\rho), & \text{if } a_1 < \rho a_2,\\ \frac{\pi}{2}, & \text{if } a_1 = \rho a_2,\\ \arctan(a_2\overline{\rho}/\rho), & \text{if } a_1 > \rho a_2.\end{cases}$$

Theorem 3.9. *For the modified SABR model* (3.1)*, the limit* (3.12) *of the mass at zero satisfies* $\mathbb{P}_\infty = \int_0^\infty \mathrm{d}t \int_0^t f(s,t)\mathrm{d}s$*, where for any* $s < t$*,*

$$f(s,t) = \frac{\pi \sin(\alpha)}{2\alpha^2 (t-s)\sqrt{s(t - s\cos^2(\alpha))}} \exp\left(-\frac{r_0^2}{2s}\frac{t - s\cos(2\alpha)}{2t - s(1+\cos(2\alpha))}\right)$$
$$\times \sum_{n=1}^{\infty} n \sin\left(\frac{n\pi(\alpha - \theta_0)}{\alpha}\right) \mathrm{I}_{\frac{n\pi}{2\alpha}}\left(\frac{r_0^2}{2s}\frac{t-s}{2t - s(1+\cos(2\alpha))}\right).$$

where I_z *denotes the modified Bessel function of the first kind of order* z *(see* [10, Page 638]*).*

Remark 3.10. Note that when $\beta = 0$, the model (3.1) exactly corresponds to the original SABR mode (1.1) with $\beta = 0$. In Theorem 3.9 above, a_1 is then equal to the starting point x_0.

Proof. Recalling the process $\widehat{X}$ in (3.7), and the SDE (3.8), we wish to apply [28, Theorem 3.1] to (3.8). Consider the system of SDEs

$$\begin{aligned} \mathrm{d}\widetilde{X}_t &= \mathrm{d}\widetilde{W}_t, & \widetilde{X}_0 &= \widehat{x}_0, \\ \mathrm{d}\widetilde{Y}_t &= \nu \mathrm{d}\widetilde{Z}_t, & \widetilde{Y}_0 &= y_0, \\ \mathrm{d}\langle \widetilde{W}, \widetilde{Z}\rangle_t &= \rho \mathrm{d}t, \end{aligned} \tag{3.13}$$

where $(\widetilde{W}, \widetilde{Z})$ is a two-dimensional standard Brownian motion. With the time-change process

$$\tau(t) := \inf\left\{u \geq 0 : \int_0^u \widetilde{Y}_s^{-2}\mathrm{d}s \geq t\right\}, \tag{3.14}$$

Theorem 3.1 in [28] implies that

$$\widehat{X}_t = \widetilde{X}_{\tau(t)} \qquad \text{and} \qquad Y_t = \widetilde{Y}_{\tau(t)}, \tag{3.15}$$

for all $t \geq 0$. In addition, the transformation (3.7) gives, for all $t \geq 0$,

$$X_t = \left(x_0^{1-\beta} + (1-\beta)\widetilde{W}_{\tau(t)}\right)^{1/(1-\beta)}.$$

Let now ε denote the explosion time of (3.13), namely the first time that either $\widetilde{X}$ or $\widetilde{Y}$ hits zero. It is also the first time that the process $\widetilde{W}$ hits the level $-\widehat{x}_0$ or that $\widetilde{Z}$ hits $-y_0/\nu$. Set

$$\Gamma_t := \int_0^t \widetilde{Y}_s^{-2}\mathrm{d}s \qquad \text{and} \qquad \zeta := \lim_{t \uparrow \zeta} \Gamma_t.$$

The process Γ is strictly increasing and continuous, so that its inverse Γ^{-1} is well defined, and clearly the time-change process (3.14) satisfies $\tau = \Gamma^{-1}$. Consider a new filtration $\mathcal{G}$ and two processes W and Z defined, for each $t \geq 0$, by $\mathcal{G}_t := \mathcal{F}_{\tau(t)}$,

$$W_t := \int_0^{\tau(t)} \frac{\mathrm{d}\widetilde{W}_s}{\widetilde{Y}_s}\mathrm{d}s \qquad \text{and} \qquad Z_t := \int_0^{\tau(t)} \frac{\mathrm{d}\widetilde{Z}_s}{\widetilde{Y}_s}\mathrm{d}s.$$

Up to time ζ, W and Z are $\mathcal{G}$-adapted Brownian motions, and the system $(W, Z, \widehat{X}, Y)$ is a weak solution to (3.8). It is therefore clear that $\mathbb{P}\left(\tau_0^{\widetilde{X}} \in \mathrm{d}s, \tau_0^{\widetilde{Y}} \in \mathrm{d}t\right) = \mathbb{P}\left(\tau_{-\widehat{x}_0}^{\widetilde{W}} \in \mathrm{d}s, \tau_{-y_0/\nu}^{\widetilde{Z}} \in \mathrm{d}t\right)$. Moreover, it follows from [38, Equation 3.2] with $\vec{\mu} = \vec{0}$, $\vec{x}_0 = (\widehat{x}_0, y_0)$, and

$$\sigma = \begin{pmatrix} \overline{\rho} & \rho \\ 0 & \nu \end{pmatrix},$$

that $\mathbb{P}\left(\tau_0^{\widetilde{X}} \in \mathrm{d}s, \tau_0^{\widetilde{Y}} \in \mathrm{d}t\right) = f(s,t)\mathrm{d}s\mathrm{d}t$, where the function f is defined in Theorem 3.9, so that

$$\mathbb{P}\left(\tau_0^{\widetilde{X}} < \tau_0^{\widetilde{Y}}\right) = \int_0^\infty \mathrm{d}t \int_0^t f(s,t)\mathrm{d}s. \tag{3.16}$$

Reversing the arguments presented in [13, 28], the probability $\mathbb{P}(\tau_0^{\widetilde{X}} < \tau_0^{\widetilde{Y}})$ coincides with the probability that the process $\widehat{X}$ hits zero over the time horizon $[0, \infty)$. Indeed, through (3.15), the time change (3.14) converts the Brownian motion $\widetilde{Y}$ into a geometric Brownian motion Y started at $y_0 > 0$, so that the (a.s. finite) point $\tau_0^{\widetilde{Y}}$ is mapped to $\tau_0^{Y} = \infty$. Therefore the time-changed process $\widetilde{X}$ over $[0, \tau_0^{\widetilde{Y}})$ corresponds to $\widehat{X}$ considered over $[0, \infty)$ and, using (3.15), we obtain

$$\mathbb{P}\left(\tau_0^{\widetilde{X}} < \tau_0^{\widetilde{Y}}\right) = \mathbb{P}\left(\tau_0^{\widehat{X}} < \tau_0^{Y}\right) = \mathbb{P}\left(\tau_0^{\widehat{X}} < \infty\right) = \mathbb{P}\left(\widehat{X}_t = 0, \text{ for some } t \in (0, \infty)\right),$$

and Theorem 3.9 follows from (3.9) and (3.16). □

Remark 3.11. For the normal SABR model ($\beta = 0$) in (1.1), Hobson [28, Example 5.2] found the following explicit formula for the price process X:

$$X_t = \frac{\rho}{\nu}\left(\widetilde{Y}_{\tau(t)} - y_0\right) + \overline{\rho}^2 \widetilde{Z}_{\tau(t)}, \qquad \text{for all } t \geq 0,$$

where the process $\widetilde{Y}$ and the Brownian motion $\widetilde{Z}$ are the same as in (3.13), and τ is defined in (3.14).

Remark 3.12. For $\beta = 1$, the SDEs of the SABR and the modified SABR models read

$$\mathrm{d}X_t = X_t\left(Y_t \mathrm{d}W_t\right), \quad X_0 = x_0, \qquad \text{and} \qquad \mathrm{d}X_t = X_t\left(Y_t \mathrm{d}W_t + \frac{1}{2}Y_t^2 \mathrm{d}t\right), \quad X_0 = x_0,$$

respectively, and, by the Doléans-Dade formula [43, Section IX-2], the solutions to these equations are exponential functionals, and therefore do not exhibit mass at the origin.

Remark 3.13. In the uncorrelated case $\rho = 0$, the expressions in Theorem 3.9 simplify to $\alpha = \frac{\pi}{2}$, $\theta_0 = \arctan\left(\frac{a_2}{a_1}\right)$, $r_0 = \sqrt{a_1^2 + a_2^2}$, and

$$f(s,t) = \frac{2}{\pi(t-s)\sqrt{st}} \exp\left(-\frac{r_0^2(t+s)}{4st}\right) \sum_{n=1}^{\infty} n \sin\left(2n\left(\frac{\pi}{2} - \theta_0\right)\right) \mathrm{I}_n\left(\frac{r_0^2(t-s)}{4st}\right).$$

4. Implied volatility and small-strike expansions

In this section, we show how the results above on the mass at zero in the SABR model can be used to infer information about the corresponding implied volatility smile. We recall that the implied volatility is simply the Black-Scholes volatility parameter that allows to match observed (or computed) European option prices. It obviously depends on strikes and maturities, and we refer the reader to [17] for more details. As noted on Page 16, one possible explanation for the inaccuracies of the 'classical' implied volatility asymptotic expansion [27] in the vicinity of zero lies in the breakdown of the commutativity of Laplace-Beltrami operators (cf. (B.1) and (B.4)), which is used in their proof, when passing from the hyperbolic heat kernel to the heat kernel on the general SABR plane. In the case where $\beta = \rho = 0$, the infinitesimal generator corresponding to the SDE (1.1) is uniformly elliptic and the heat kernel is known. We plot below (Figure 4) the implied volatility expansion derived in [41]–a slightly refined version of the one in [27]–and highlights the fact that it can yield arbitrage. As explained in [18], the density of the log stock price $\log(X)$ (or log forward rate) can be expressed directly in terms of the implied volatility (see [18, Proof of Lemma 2.2]), and negative densities obviously yield arbitrage opportunities.

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 [35]:

$$\limsup_{K \downarrow 0} \frac{I_T(K)}{\sqrt{|\log K|}} = \sqrt{\frac{2}{T}}. \tag{4.1}$$

This behaviour was recently refined by de Marco et. al. and Gulisashvili independently [12, 22]. Assuming that there exists $\varepsilon > 0$ such that $\mathbb{P}(X_T \leq K) - \mathbb{P}(X_T = 0) = \mathcal{O}(\varepsilon)$ as K tends to zero, de Marco et al. [12] derive the following small-strike asymptotic formula:

$$I_T(K) = \sqrt{\frac{2|\log K|}{T}} + \frac{\mathcal{N}^{-1}(\mathrm{m}_T)}{\sqrt{T}} + \frac{(\mathcal{N}^{-1}(\mathrm{m}_T))^2}{2\sqrt{2T|\log K|}} + \Phi(K), \tag{4.2}$$

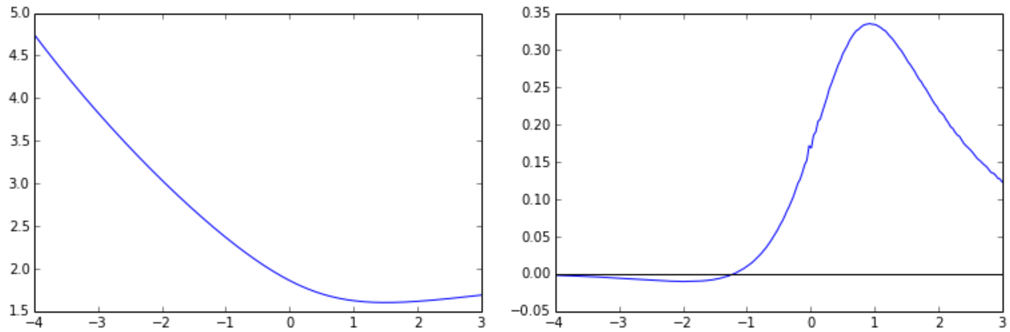

FIGURE 3. Density (right) of the log process $\log(X)$ obtained from the implied volatility expansion [41] (left) with $(\nu, \beta, \rho, x_0, y_0, T) = (0, 1, 0.6, 0.05, 0.5, 1.2)$. The mass at zero, computed using (2.2) is equal to 4.5%.

where $\mathrm{m}_T := \mathbb{P}(X_T = 0)$ is the mass at the origin, $\mathcal{N}$ the Gaussian cumulative distribution function, and $\Phi : (-\infty, 0) \to \mathbb{R}$ a function satisfying $\limsup_{K\downarrow 0} \sqrt{2T|\log K|}|\Phi(K)| \leq 1$. Gulisashvili [22] obtained an alternative formulation (removing the assumption on the decay of the probability of X_T near zero); however, since we only wish here to highlight the inaccuracy of Hagan's (or Obłój's) expansion in low-strike regimes, we omit a precise formulation of his result and refer the interested reader to this paper for full details. In Figure 4 below, we visually quantify how 'wrong' Hagan's expansion is for small strikes in the presence of a mass at the origin. We plot the functions $k \in \mathbb{R} \mapsto I_T(\mathrm{e}^k \sqrt{T/|k|})$ which, from (4.2) has to be bounded by $\sqrt{2}$ in order to avoid arbitrage, and compare it to the first and second order of (4.2), using (2.10) to compute the (large-time) mass at zero. We consider two parameter sets, one for which the large-time mass is small, and the other which yields a large mass at the origin. As the mass becomes small, Hagan's (or Obłój's) approximation becomes more accurate. This holds in particular as the parameter β gets close to one, as indicated in Section 2.3.1 above. In the limit as $\beta = 1$, the mass becomes null, and Hagan's expansion, as noted in the literature indeed becomes arbitrage free.

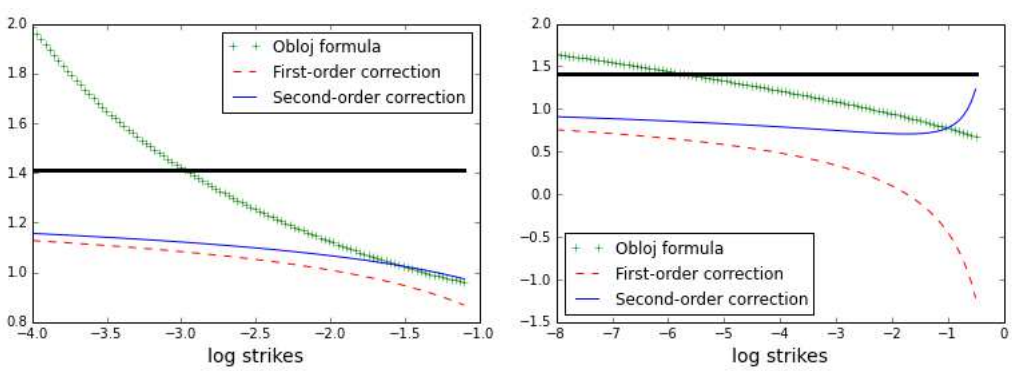

FIGURE 4. The black line marks the level $\sqrt{2}$. The parameters are $(\nu, \beta, \rho, x_0, y_0, T) = (0.3, 0, 0, 0.35, 0.05, 10)$ for the left plot, and $(\nu, \beta, \rho, x_0, y_0, T) = (0.6, 0.6, 0, 0.08, 0.015, 10)$ for the right graph. Obłój's implied volatility expansion clearly violates 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.

Appendix A. Proofs of Section 2

A.1. **Proof of Proposition 2.5.** We adapt here Gerhold's proof in [19] to our case, 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

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

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

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

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:

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

The first one is standard [39, Section 3.5]. As for the second one, the representation (2.7) yields

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

where

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

for $k \geq 0$. Now, clearly $|\gamma_k| \leq 2^{-k}$ and $\gamma_k \sim 2^{-k}$ as v tends to infinity. Note furthermore that from [1, Formula 13.6.3], we have

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

where again I_v denotes the modified Bessel function of the first kind [10, Page 638], so that, using (A.2) and [19, Equation 9], we have, uniformly in v,

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

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

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

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

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

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

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

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 [19]. 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 [19], we can therefore deform the contour of integration in (A.1) around the saddlepoint u_y to obtain, as y tends to zero:

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

Let λ denote the real integration variable, so that $u = u_y + \mathrm{i}\lambda$. Around the saddlepoint ($\lambda = 0$), we have the uniform Taylor series expansions:

$$\sqrt{u} = \sqrt{u_y} + \frac{\mathrm{i}\lambda}{2\sqrt{u_y}} + \frac{\lambda^2}{8u_y^{3/2}} + \mathcal{O}\left(\frac{\lambda^3}{u_y^{5/2}}\right), \qquad \log u = \log u_y + \frac{\mathrm{i}\lambda}{u_y} + \frac{\lambda^2}{2u_y^2} + \mathcal{O}\left(\frac{\lambda^3}{u_y^3}\right),$$

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

so that

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

where the coefficients in front of λ cancelled out from the saddlepoint equation, 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 2.5. Note that by bootstrapping (see Section A.1.1 for details), the expansion

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

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

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

Now,

$$\int_{-h}^{h} \mathrm{e}^{-M_y y^2}\mathrm{d}y = \frac{1}{\sqrt{2M_y}}\int_{-h\sqrt{2M}}^{h\sqrt{2M_y}} \mathrm{e}^{-\omega^2/2}\mathrm{d}\omega \sim \frac{1}{\sqrt{2M_y}}\int_{\mathbb{R}} \mathrm{e}^{-\omega^2/2}\mathrm{d}\omega = \sqrt{\frac{\pi}{M_y}} \sim \frac{\sqrt{\pi}|\log(y)|}{y^{3/2}},$$

and we can therefore write (A.5) as

(A.9)
$$
\begin{aligned}
m_y(\mu,z) &\sim \frac{\sqrt{z}}{2\mathrm{i}\pi}\int_{u_y-\mathrm{i}h}^{u_y+\mathrm{i}h} \mathrm{e}^{\psi_y(u)}\mathrm{d}u\\
&\sim \frac{\sqrt{z}}{2\pi}\exp\left[u_y y+\sqrt{u_y}\left(\alpha-\frac{\log(u_y)}{2}\right)-\frac{\log(u_y)}{4}+\frac{\mu\log(u_y)}{2}\right]\int_{-h}^{h}\mathrm{e}^{-M_y\lambda^2}\mathrm{d}\lambda\\
&\sim \frac{\sqrt{z}}{2\pi}\exp\left[u_y y+\sqrt{u_y}\left(\alpha-\frac{\log(u_y)}{2}\right)-\frac{\log(u_y)}{4}+\frac{\mu\log(u_y)}{2}\right]\frac{\sqrt{\pi}|\log(y)|}{y^{3/2}}\\
&= \frac{\sqrt{z}}{2\pi}\exp\left[\left(\frac{1}{2}-\mu\right)+\frac{\log(u_y)}{2}\left(\mu-\frac{1}{2}\right)-u_y y+\sqrt{u_y}\right]\frac{\sqrt{\pi}|\log(y)|}{y^{3/2}}\\
&= \frac{\sqrt{z}}{2\sqrt{\pi}}\exp\left(\frac{1}{2}-\mu\right)\frac{|\log(y)|}{y^{3/2}}u_y^{\frac{1}{2}(\mu-\frac{1}{2})}\exp\left(-u_y y+\sqrt{u_y}\right)\\
&= \frac{\sqrt{z}}{2\sqrt{\pi}}|\log(y)|\exp\left[-\frac{\log(y)^2}{4y}+\frac{1}{2}\frac{|\log(y)|}{y}+\left(\frac{1}{2}-\mu\right)\left(1-\frac{1}{2}\log\left(\frac{\log(y)^2}{4y^2}\right)\right)\right]\left(\frac{1}{y^{3/2}}+\mathcal{O}\left(y^{3/2}\right)\right),
\end{aligned}
$$

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 [19, 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 $\int_{u_y-\mathrm{i}\infty}^{u_y+\mathrm{i}\infty}\exp(\psi_y(u))\mathrm{d}u$, 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 $\int_{u_y+\mathrm{i}h}^{u_y+\mathrm{i}\infty}\mathrm{e}^{\psi_y(u)}\mathrm{d}u$. The analysis is then split into looking at the inner tail $h\leq\lambda<\mathrm{e}^{\log(1/t)^2/4}$ and at the outer tail $\lambda\geq\mathrm{e}^{\log(1/t)^2/4}$. Similarly to [19, Equation (10)], the estimate

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

then prevails for the outer tail. Furthermore, for any real number B, [19, 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 $\mathrm{e}^{\log(1/t)^2/4}$; 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 [19, Table 1], and the total (both tails) error resulting from the completion to Gaussian integral

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

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

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

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 $\frac{1-2\mu}{8u_y^2}$ in the definition (2.9) of M_y is of higher order, therefore we can henceforth 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 $\widetilde{u} \equiv u_y/2$, the approximation of the saddlepoint simplifies to

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

Thus M_y is up to constants of the same form as [19, 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 (2.8)

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

when setting $c(u) \equiv \left(\frac{\log(\sqrt{u})}{\sqrt{2}} - \frac{\rho}{\sqrt{2}} + \frac{k}{4\sqrt{u}}\right)$, $\rho = \log\left(\frac{z}{\sqrt{2}}\right)$ and $k := 4\mu - 2$, and with $u_0 = 2u_y$ becomes $\sqrt{u_0} \equiv y^{-1}c(u_0)$, where

$$\frac{\log(\sqrt{u_0})}{\sqrt{2}} = \frac{\log\left(c\left(u_0\right) + \log\left(\frac{1}{y}\right)\right)}{\sqrt{2}}.$$

Hence, bootstrapping as in [19] 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 $\log(1/y)$

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

and using the fact that both

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

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

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

so that, by bootstrapping we also recover the form of [19, Equation (13)], at $\widetilde{u} = \frac{1}{2}u_y$:

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

Appendix B. Reminder on the heat equation on manifolds

Proposition B.1. *Let $k \in \mathbb{N} \cup \{\infty\}$ be an arbitrary integer, M_1, M_2 two C^{k+2}-manifolds and $\phi : M_2 \to M_1$ a C^{k+2}-diffeomorphism which is an isometry between (M_2, g_2) and (M_1, g_1). The Laplace-Beltrami operator Δ_{g_i}, $i \in \{1, 2\}$ commutes with ϕ in the sense that the equation*

$$\Delta_{g_2}(\phi^*(f)) = \phi^*(\Delta_{g_1}(f)) \tag{B.1}$$

holds for any $f \in C^{k+2}(M_1)$.

Proof. A proof of this statement is given for example in [21, Lemma 3.27]. □

Definition B.2. Let (M, g) be a smooth Riemannian manifold and $Z \in M$. The smooth function $p_Z : (0, \infty) \times M \to \mathbb{R}$ is a fundamental solution at Z of the heat equation on (M, g) if it satisfies the following conditions:

(i) p_Z solves the heat equation on (M, g): $\Delta_g p_Z = \partial_t p_Z$, where Δ_g denotes the Laplace-Beltrami operator on (M, g);

(ii) $\lim_{t\downarrow 0} p_Z(t, \cdot) = \delta_Z$, where δ_Z denotes the Dirac measure at $Z \in M$:
$$\lim_{t\downarrow 0} \int_M p_Z(t, z) f(z) \mu_g(\mathrm{d}z) = f(Z),$$
for all test functions $f \in C_0^\infty(M)$, where $\mu_g(\mathrm{d}z)$ stands for the Riemannian volume element at $z \in M$.

The fundamental function p_Z is said to be regular if furthermore $p_Z \geq 0$ and $\int_M p_Z(t, z)\mu_g(\mathrm{d}z) \leq 1$.

Proposition B.3. *Let $k \in \mathbb{N}_0 \cup \{\infty\}$, $\phi : (M_2, g_2) \longrightarrow (M_1, g_1)$ a C^{k+2}-smooth isometry, and $p_{Z_1}^{g_1}$ the fundamental solution at Z_1 of the heat equation on (M_1, g_1). Furthermore, let $Z_2 \in M_2$ be such that $\phi(Z_2) = Z_1$. Then the map $(t, z) \mapsto p_{\phi(Z_2)}^{g_1}(t, \phi(z))$ is the (unique) fundamental solution at Z_2 of the heat equation on (M_2, g_2).*

Proof. Property (i) in Definition B.2 holds for the above map, which follows from Proposition B.1 and especially from (B.1). The operator Δ_{g_2} acts only on the space variable $\widetilde{z} \in M_2$ and not on $\widetilde{Z} \in M_2$, so that

$$\Delta_{g_2}\left(p_{\phi(Z_2)}^{g_1}(t, \phi(z))\right) = \phi^*\left(\Delta_{g_1}(p_{Z_2}^{g_1}(t, z))\right) = \phi^*\left(\frac{\partial}{\partial t}(p_{Z_2}^{g_1}(t, z))\right) = \frac{\partial}{\partial t}\left(p_{\phi(Z_2)}^{g_1}(t, \phi(z))\right), \tag{B.2}$$

where the first equality follows from (B.1). Property (ii) of Definition B.2 is a consequence of the substitution rule. Let $\widetilde{f} \in C_0^\infty(M_1)$ be a test function and $f := \phi^*\widetilde{f}$. Set $\widetilde{z} = \phi(z)$ and $\widetilde{Z} = \phi(Z)$ for any $z, Z_2 \in M_2$. Given that ϕ is an isometry, so is ϕ^{-1} and the pullback $(\phi^{-1})^*\mu_{g_2}(\mathrm{d}\cdot)$ coincides with the volume form on (M_1, g_1). Then

$$\begin{aligned}
\lim_{t\downarrow 0}\int_{M_2} p_{\phi(Z_2)}^{g_1}(t, \phi(z)) f(z)\mu_{g_2}(\mathrm{d}z) &= \lim_{t\downarrow 0}\int_{M_2} p_{\phi(Z_2)}^{g_1}(t, \phi(z))\widetilde{f}(\phi(z))\mu_{g_2}(\mathrm{d}z) \\
&= \lim_{t\downarrow 0}\int_{M_1} p_{\phi(Z_2)}^{g_1}(t, \widetilde{z})\widetilde{f}(\widetilde{z})\left((\phi^{-1})^*\mu_{g_2}\right)(\mathrm{d}\widetilde{z}) \\
&= \lim_{t\downarrow 0}\int_{M_1} p_{\phi(Z_2)}^{g_1}(t, \widetilde{z})\widetilde{f}(\widetilde{z})\mu_{g_1}(\mathrm{d}\widetilde{z}) = \widetilde{f}(\phi(Z_2)) = f(Z).
\end{aligned}$$

□

Remark B.4. Note that the fundamental solutions in Proposition B.3 are denoted with respect to the Riemannian volume form. In terms of integration with respect to the Lebesgue measure we make a slight modification of the above statement.

Let the Riemannian volume form be given in orthogonal coordinates, and let K^u and K^g denote the fundamental solutions (in terms of Lebesgue) of the heat equations corresponding to the Riemannian metrics u and g in the sense that the Radon-Nikodym derivatives with respect to the Lebesgue measure are already incorporated into the expression for K^u and K^g: if $p_Z^g(s,z)$ (resp. $p_Z^u(s,z)$) denote the fundamental solutions as in Proposition B.3, then, for any test function f,

$$\int_{\mathcal{S}} f(z) K_Z^g(s,z)\mathrm{d}z = \int_{\mathcal{S}} f(z) p_Z^g(s,z)\frac{\mathrm{d}z}{\mu_g(\mathrm{d}z)}\mu_g(\mathrm{d}z) = \int_{\mathcal{S}} f(z) p_Z^g(s,z)\frac{\mu_g(\mathrm{d}z)}{\sqrt{\det(g)}}.$$

The following lemma follows directly from Proposition B.3.

Lemma B.5. *Let K^{g_1} and K^{g_2} denote the fundamental solutions (in terms of Lebesgue) of the heat equations corresponding to the metrics g_1 and g_2:*

$$\begin{cases} \dfrac{\partial K^{g_1}}{\partial s} = \dfrac{1}{2}\Delta_{g_1} K^{g_1}, \\ K_Z^{g_1}(0,z) = \delta(z-Z), \end{cases} \quad \text{and} \quad \begin{cases} \dfrac{\partial K^{g_2}}{\partial s} = \dfrac{1}{2}\Delta_{g_2} K^{g_2}, \\ K_{\widetilde{Z}}^{g_2}(0,\widetilde{z}) = \delta(\widetilde{z}-\widetilde{Z}). \end{cases} \tag{B.3}$$

If ϕ is an isometry, then

$$K_Z^{g_1}(s,z) = \det(\nabla\phi(Z))\ K_{\phi(Z)}^{g_2}(s,\phi(z)). \tag{B.4}$$

The generators of the Brownian motions on $(\mathcal{S}, g)$ resp $(\mathbb{U}, u)$ (defined in Section 3.1) are defined on their respective spaces with $\{x \neq 0\}$ and $\{\bar{x} \neq 0\}$ for $\beta \neq 0$ respectively and read

$$\begin{aligned} \Delta_g f &= y^2\left(\beta x^{2\beta-1}\frac{\partial f}{\partial x} + x^{2\beta}\frac{\partial^2 f}{\partial x^2} + 2\rho x^\beta \frac{\partial}{\partial x}\frac{\partial f}{\partial y} + \frac{\partial^2 f}{\partial y^2}\right), \quad &\text{for any } f \in C^{k+2}(\mathcal{S}), \\ \Delta_u f &= \bar{y}^2\left(\beta \bar{x}^{2\beta-1}\frac{\partial f}{\partial \bar{x}} + \bar{x}^{2\beta}\frac{\partial^2 f}{\partial \bar{x}^2} + \frac{\partial^2 f}{\partial \bar{y}^2}\right), \quad &\text{for any } f \in C^{k+2}(\mathbb{U}), \end{aligned} \tag{B.5}$$

while the infinitesimal generators of the original SABR model (1.1) are

$$\begin{aligned} \mathcal{A} f &= y^2\left(x^{2\beta}\frac{\partial^2 f}{\partial x^2} + 2\rho x^\beta \frac{\partial}{\partial x}\frac{\partial f}{\partial y} + \frac{\partial^2 f}{\partial \bar{y}^2}\right), \quad &\text{for any } f \in C^{k+2}(\mathcal{S}), \\ \mathcal{A}_{\rho=0} f &= \bar{y}^2\left(\bar{x}^{2\beta}\frac{\partial^2 f}{\partial \bar{x}^2} + \frac{\partial^2 f}{\partial \bar{y}^2}\right), \quad &\text{for any } f \in C^{k+2}(\mathbb{U}), \end{aligned} \tag{B.6}$$

Note that for $\beta = 0$ the operators Δ_g and $\mathcal{A}$ (resp. Δ_u and $\mathcal{A}_{\rho=0}$) coincide.

References

[1] M. Abramowitz, I. A. Stegun. Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables. Dover Publications, 1964.

[2] L. Alili and J.-C. Gruet. An explanation of a generalized Bougerol's identity in terms of hyperbolic Brownian motion. *Exponential Functionals and Principal Values related to Brownian Motion*. Revista Matematica Iberoamericana. MR1648653, 1997.

[3] J. Andreasen and B. Huge. ZABR – Expansions for the Masses. Preprint, SSRN//1980726, 2011.

[4] A. Antonov, M. Konikov and M. Spector. The Free Boundary SABR: Natural Extension to Negative Rates. Preprint, SSRN//2557046, 2015.

[5] A. Antonov and M. Spector. Advanced Analytics for the SABR Model. Preprint, SSRN//2026350, 2012.

[6] P. Balland and Q. Tran. SABR Goes Normal. *Risk*, June issue: 76-81, 2013.

[7] 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.

[8] S. Benaim, P. Friz. Regular variation and smile asymptotics. *Mathematical Finance*, 19: 1-12, 2009.

[9] J. Blath, L. Döring and A. Etheridge, On the moments and the interface of the symbiotic branching model. *The Annals of Probability*, 39: 252-290, 2011.

[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. Efficient unbiased simulation scheme for the SABR stochastic volatility model. Preprint, http://ta.twi.tudelft.nl/mf/users/oosterle/oosterlee/SABRMC.pdf, 2011.

[12] S. De Marco, C. Hillairet, and A. Jacquier. Shapes of implied volatility with positive mass at zero. Preprint, arXiv:1310.1020, 2013.

[13] L. Döring, and B. Horvath, Time-change techniques for SABR-type models. Working paper, 2015.
[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] M. Forde, H. Zhang. Sharp tail estimates for the correlated SABR model. *Preprint*, 2014.
[17] J. Gatheral. The Volatility Surface: a Practitioner's Guide. Wiley, 2006.
[18] J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1): 59-71, 2014.
[19] S. Gerhold. The Hartman-Watson distribution revisited: asymptotics for pricing Asian options. *Journal of Applied Probability*, 48(3): 892-899, 2011.
[20] I. S. Gradshteyn and I. M. Ryzhik. Table of Integrals, Series, and Products, Sixth Edition. Alan Jeffrey and Daniel Zwillinger (eds.) Academic Press, 2000.
[21] A. Grigor'yan. Heat Kernel and Analysis on Manifolds. American Mathematical Society, 2009.
[22] A. Gulisashvili. Left-wing asymptotics of the implied volatility in the presence of atoms. *International Journal of Theoretical and Applied Finance*, 2015.
[23] A. Gulisashvili. Analytically Tractable Stochastic Stock Price Models. Springer, 2012.
[24] 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.
[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, Volume 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. Preprint, SSRN//1489428, 2009.
[30] S. Iyengar. Hitting lines with two-dimensional Brownian motion. *SIAM App. Math.*, 45: 983-989, 1985.
[31] J. Jacod and A.N. Shiryaev. Limit Theorems for Stochastic Processes, 2nd Edition. Springer Berlin, 2003.
[32] M. Jeanblanc, M. Yor, and M. Chesney. Mathematical Methods for Financial Markets. Springer Finance, 2009.
[33] B. Jourdain. Loss of martingality in asset price models with lognormal stochastic volatility. *International Journal of Theoretical and Applied Finance*, 13: 767-787, 2004.
[34] R. Rebonato. A simple approximation for the no-arbitrage drifts for LMM-SABR-family interest-rate models. Preprint, papers.ssrn.com/sol3/papers.cfm?astract-id=2560241, 2014.
[35] R. W. Lee. The moment formula for implied volatility at extreme strikes. *Math Finance*, 14: 469-480, 2004.
[36] H. Matsumoto and M. Yor. Exponential functionals of Brownian motion, I: Probability laws at fixed time. *Probability Surveyes*, 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] A. Metzler. On the first passage problem for correlated Brownian motion. *Statistics and Probability Letters*, 80: 277-284, 2010.
[39] P. D. Miller. Applied Asymptotic Analysis. Graduate Studies in Mathematics, volume 75, American Mathematical Society, 2006.
[40] H.P. McKean. An upper bound to the spectrum of on a manifold of negative curvature. *Journal of Differential Geometry*, 4: 359-366, 1970.
[41] J. Obłój. Fine-tune your Smile: Correction to Hagan et al. *Wilmott Magazine*, May issue, 2008.
[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, Volume 110, 2015.
[43] D. Revuz and M. Yor. Continuous Martingales and Brownian Motion. Springer, Berlin, 2004.
[44] N. A. Sidorova and O. Wittich. Construction of surface measures for Brownian motion. In Trends in Stochastic analysis - a Festschrift in honour of Heinrich v. Weizscker, 2009.
[45] A. Veraart and M. Winkel. Time change. Encycl. of Quant. Fin. (Editor R. Cont), Wiley, IV: 1812-1816, 2010.
[46] V. Volkonskii. Random time changes in strong Markov processes. *Theory of Prob. and App.*, 3: 310-326, 1958.
[47] 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

Departement of Mathematics, ETH Zürich
E-mail address: blanka.horvath@math.ethz.ch

Department of Mathematics, Imperial College London
E-mail address: a.jacquier@imperial.ac.uk