risks

MDPI

Article

Probability Density of Lognormal Fractional SABR Model †

Jiro Akahori [1], Xiaoming Song [2] and Tai-Ho Wang [1,3,*]

1 Department of Mathematical Sciences, Ritsumeikan University, Noji-Higashi 1-1-1, Kusatsu 525-8577, Shiga, Japan; akahori@se.ritsumei.ac.jp
2 Department of Mathematics, Drexel University, 32nd and Market Streets, Philadelphia, PA 19096, USA; song@math.drexel.edu
3 Department of Mathematics, Baruch College, The City University of New York, 1 Bernard Baruch Way, New York, NY 10010, USA
* Correspondence: tai-ho.wang@baruch.cuny.edu

† In memory of Peter Carr, dear friend and inspirational financial engineer.

Abstract: Instantaneous volatility of logarithmic return in the lognormal fractional SABR model is driven by the exponentiation of a correlated fractional Brownian motion. Due to the mixed nature of driving Brownian and fractional Brownian motions, probability density for such a model is less studied in the literature. We show in this paper a bridge representation for the joint density of the lognormal fractional SABR model in a Fourier space. Evaluating the bridge representation along a properly chosen deterministic path yields a small time asymptotic expansion to the leading order for the probability density of the fractional SABR model. A direct generalization of the representation of joint density often leads to a heuristic derivation of the large deviations principle for joint density in a small time. Approximation of implied volatility is readily obtained by applying the Laplace asymptotic formula to the call or put prices and comparing coefficients.

Keywords: asymptotic expansion; lognormal fractional SABR model; mixed fractional Brownian motion; Malliavin calculus; bridge representation

check for updates

Citation: Akahori, Jiro, Xiaoming Song, and Tai-Ho Wang. 2022. Probability Density of Lognormal Fractional SABR Model. *Risks* 10: 156. https://doi.org/10.3390/risks10080156

Academic Editors: Dan Pirjol and Lingjiong Zhu

Received: 27 June 2022
Accepted: 20 July 2022
Published: 2 August 2022

1. Introduction

The celebrated Black and Black-Scholes-Merton models have been the benchmark for European options on currency exchange, interest rates, and equities since the inauguration of the trading on financial derivatives. However, empirical evidence has shown that the main drawback of these models is the assumption of constant volatility; the key parameter required in the calculation of option premia under such models. The volatility parameters induced from market data are in fact nonconstant across markets; dubbed as *volatility smile*.

The Stochastic $\alpha\beta\rho$ (SABR hereafter) model, suggested by Hagan, Lesniewski, and Woodward in Hagan et al. (2015), is one of the models, such as local volatility models, stochastic volatility models, and exponential Lévy type of models, etc, that attempts to capture the volatility smile effect. Furthermore, as opposed to local volatility models, in the SABR model the volatility smile moves in the same direction as the underlying with time, see Hagan et al. (2002).

The SABR model is depicted by the following system of stochastic differential equations (SDEs):

$$dF_t = \alpha_t F_t^\beta dW_t, \quad F_0 = F, \tag{1}$$

$$d\alpha_t = \nu \alpha_t dZ_t, \quad \alpha_0 = \alpha, \tag{2}$$

with $\beta \in [0,1]$, where F_t denotes the forward price and α_t the instantaneous volatility. W_t and Z_t are correlated Brownian motions with a constant correlation coefficient ρ. The SABR model is at times referred to as lognormal SABR when $\beta = 1$. The SABR formula is an asymptotic expansion for the implied volatilities of call options with various strikes with small expiry times. For the reader's convenience, we reproduce the SABR formula in the following. Let $\sigma_{BS}(K,\tau)$ be the implied volatility of a vanilla option struck at K and time to expiry τ. The SABR formula states

$$\sigma_{BS}(K,\tau) = \nu \frac{\log(F/K)}{D(\zeta)}\{1 + O(\tau)\} \tag{3}$$

as the time to expiry τ approaches 0. The function D and the parameter ζ involved in (3) are defined respectively as

$$D(\zeta) = \log\left(\frac{\sqrt{1 - 2\rho\zeta + \zeta^2} + \zeta - \rho}{1 - \rho}\right)$$

and

$$\zeta = \begin{cases} \frac{\nu}{\alpha}\frac{F^{1-\beta} - K^{1-\beta}}{1-\beta} & \text{if } \beta \neq 1; \\ \frac{\nu}{\alpha}\log\left(\frac{F}{K}\right) & \text{if } \beta = 1. \end{cases}$$

Generally, the SABR formula is given one order higher, up to order τ. Here we present only the zeroth order for our own purpose.

The geometry of the SABR model is isometrically diffeomorphic to the two-dimensional hyperbolic space, also known as the Poincaré plane. This isometry leads to a derivation of the SABR Formula (3) based on an expression of the heat kernel, known as the McKean kernel, on Poincaré plane. In particular, the lowest order term in (3) has a geometric interpretation. The function D is the geodesic distance from the spot value (F_0, α_0) to the vertical line $F = K$ in the upper half plane $\{(F,\alpha) \in \mathbb{R}^2 : \alpha \geq 0\}$. Hence, the lowest order term in (3) is indeed the ratio between the absolute value of logmoneyness, i.e., $\log(K/F_0)$, and the geodesic distance from (F_0, α_0) to the vertical line $F = K$ in the upper half-plane. We refer readers interested in this topic to Hagan et al. (2015) for more detailed discussions. As expression for heat kernel on hyperbolic space is concerned, Ikeda and Matsumoto in Ikeda and Matsumoto (1999) provided a probabilistic approach and obtained, among other interesting results, a representation for the transition density of hyperbolic Brownian motion, i.e., the heat kernel over the Poincaré plane. See Theorem 2.1 in Ikeda and Matsumoto (1999) for details.

The aforementioned nice isometry between the SABR model and Poincaré plane breaks down if the volatility process, i.e., the α_t process in (1), is driven by a fractional Brownian motion such as the second equation in (6) considered in the paper. Moreover, due to the lack of Markovianity of fractional Brownian motions, thus the nonexistence of the forward and backward Kolmogorov equations, the classical asymptotic expansion approaches, such as the heat kernel or WKB expansion, are no longer applicable. In this regard, the probabilistic approach in Ikeda and Matsumoto (1999) is more applicable and tractable when dealing with processes driven by fractional Brownian motions.

The volatility process is generally conceived as behaving "fractionally" in that the driving noise is a fractional process, e.g., a fractional Brownian motion with a Hurst exponent other than a half. For a far from an exhaustive list, models that attempt to incorporate the fractional feature of volatility include: the ARFIMA model in Granger and Joyeux (1980) and the FIGARCH model Baillie et al. (1996) for discrete-time models; the long memory stochastic volatility model in Comte and Renault (1998) and the affine fractional stochastic volatility model in Comte et al. (2012) for continuous time models. Somewhat on the contrary, in a recent study in Gatheral et al. (2018), the Hurst exponent H is estimated as being less than a half; thereby indicating antipersistency as opposed to the persistency of the volatility process. For a more detailed and in-depth consideration of this issue, we refer interested readers to the discussions in Cont and Das (2022) and Rogers (2019). It is also worth mentioning that generalizations of the Heston model to the fractional version have been considered in El Euch and Rosenbaum (2019) and Guennoun et al. (2018). Heston-related models are usually dealt with via the characteristic or moment-generating functions. However, in this paper, we take the approach following closely the methodology in Ikeda and Matsumoto (1999). As arbitrage in the modeling is concerned, we remark that, in contrast with the models discussed in for instance Jarrow et al. (2009) and Mishura (2008) within which the underlying prices were assumed driven by fractional Brownian motions, the model considered in the paper is free of arbitrage opportunity since it is the volatility process that is driven by a fractional Brownian motion while the underlying itself is still driven by a (correlated) Brownian motion.

In order to embed the empirically observed fractional feature of the volatility process into the classical SABR model, we suggest in this paper a fractional version of the SABR model as in (6). Modulo a mean-reversion component, this model aligns with the model statistically tested in Gatheral et al. (2018). The main observation in Gatheral et al. (2018) is that in using the square root of the realized variance as a proxy for the instantaneous volatility, the logarithm of the volatility process behaves like a fractional Brownian motion in almost any time scale of frequency. The Hurst exponent H inferred from the time series data is less than a half; indeed, $H \approx 0.1$, see also Cont and Das (2022) and Rogers (2019). This observation of a small Hurst exponent in the volatility process analyzes the model as more technical and challenging from a stochastic analysis point of view. To our knowledge, most of the small time asymptotic expansions for processes driven by fractional Brownian motions have restrictions on the Hurst exponent H of the driving fractional Brownian motion, mostly $H \geq \frac{1}{4}$. One of the advantages of the approach undertaken in the current paper is that it works without restriction on the Hurst exponent H. The key ingredient is a representation in a Fourier space, which we call the bridge representation in Section 2, for the joint density of log spot and volatility, see (9).

A small time asymptotic expansion of the joint density is readily obtained from the bridge representation. The idea is to approximate the conditional expectation in the bridge representation by a judiciously chosen deterministic path since, conditioned on the initial and terminal points, at each point in time a Gaussian process will not wander too far away from its expectation. As long as an asymptotic expansion for the density of the underlying asset is available, obtaining an expansion for implied volatility is almost straightforward by basically comparing the coefficients with a similar expansion obtained by using the lognormal density on the Black or the Black-Scholes-Merton side.

The methodology of deriving the bridge representation (9) can be generalized directly to obtain a bridge representation for the joint density multiple times; hence inducing a representation for finite-dimensional distributions of the fractional SABR model, see Theorem 4. Based on this bridge representation for finite-dimensional distributions, Section 5 is devoted to a heuristic yet appealing derivation of the large deviations principle for the joint density of the fractional SABR model in small time. This large deviations principle in a sense can be regarded as defining a "geodesic distance" over the fractional SABR plane since, as we shall show in Section 5, it recovers the energy functional on the Poincaré plane when $H = \frac{1}{2}$. We leave the rigorous proof of the large deviations principle in future work. An immediate consequence of this large deviation principle is the fractional SABR formula (to the lowest order) (26) which recovers the classical SABR formula when $H = \frac{1}{2}$. The fractional SABR Formula (26) pertains to the guiding principle that the lowest order term in the implied volatility expansion is given by the ratio between the absolute value of the logmoneyness and the geodesic distance to the vertical line $F = K$.

The rest of the paper is organized as follows. The fractional SABR model is specified and the bridge representation for joint density is shown in Section 2. Sections 3 and 4 provide small time asymptotic expansions of the joint density and of the implied volatilities respectively. Section 5 presents the bridge representation for finite-dimensional distributions and the large deviations principle. Finally, the paper concludes in Section 6 with discussions.

2. Model Specification

Throughout the text, $B = \{B_t, t \geq 0\}$ and $W = \{W_t, t \geq 0\}$ denote two independent standard Brownian motions defined on the filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ satisfying the usual conditions. Let $B^H = \{B_t^H, t \geq 0\}$ be a fractional Brownian motion with Hurst exponent $H \in (0,1)$ generated by B (see Decreusefond and Üstünel 1999), i.e.,

$$B_t^H = \int_0^t K_H(t,s) dB_s,$$

where K_H is the Molchan-Golosov kernel

$$K_H(t,s) = c_H (t-s)^{H-\frac{1}{2}} F\left(H - \frac{1}{2}, \frac{1}{2} - H, H + \frac{1}{2}; 1 - \frac{t}{s}\right) \mathbf{1}_{[0,t]}(s), \tag{4}$$

with $c_H = \left[\frac{2H\Gamma\left(\frac{3}{2}-H\right)}{\Gamma(2-2H)\Gamma\left(H+\frac{1}{2}\right)}\right]^{1/2}$ and F is the Gauss hypergeometric function. Also, the autocovariance function of a fractional Brownian motion is denoted by $R(t,s)$ and defined as

$$R(t,s) = \mathbb{E}(B_t^H B_s^H) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right). \tag{5}$$

Lastly, we assume that all random variables and stochastic processes are defined on $(\Omega, \mathcal{F}_t, \mathbb{P})$.

2.1. The Model

We study the following lognormal fractional SABR (fSABR hereafter) model in risk-neutral probability (for simplicity, interest and dividend rates are both assumed zero):

$$\begin{cases} S_t = s_0 + \int_0^t \alpha_r S_r (\rho dB_r + \bar{\rho} dW_r), \\ \\ \alpha_t = \alpha_0 e^{\nu B_t^H}, \end{cases} \tag{6}$$

where s_0 and α_0 are the given time zero (current observed) values for the processes S and α respectively, $\rho \in (-1,1)$ and $\bar{\rho} = \sqrt{1-\rho^2}$.

In other words, the underlying price S_t follows a stochastic volatility model with the (instantaneous) volatility process α_t, and α_t is given by the exponentiation of a correlated fractional Brownian motion. The main purpose of this section is to derive the bridge representations (9) and (13) for the joint densities of (S_t, α_t). The bridge representation is the crucial starting line in obtaining expansions and approximations of the joint densities to be discussed in Section 3.

By making a change in variables

$$X_t = \ln S_t, \qquad Y_t = \alpha_t,$$

the system (6) can be written more explicitly as

$$\begin{cases} X_t = x_0 + y_0 \int_0^t e^{\nu B_s^H} (\rho dB_s + \bar{\rho} dW_s) - \frac{y_0^2}{2} \int_0^t e^{2\nu B_s^H} ds, \\ \\ Y_t = y_0 e^{\nu B_t^H}, \end{cases} \tag{7}$$

where $x_0 = \ln s_0$ and $y_0 = \alpha_0$.

2.2. Malliavin Calculus with Respect to Brownian Motion

We provide some preliminaries on Malliavin calculus with respect to the two Brownian motions B and W in this subsection. We refer the reader to Hu (2017) and Nualart (2006) for more details.

For any fixed $T > 0$, let $\mathbf{H} = L^2([0,T])$ be the separable Hilbert space of all square-integrable real-valued functions on the interval $[0,T]$ with scalar product denoted by $\langle \cdot, \cdot \rangle_{\mathbf{H}}$. The norm of an element $h \in \mathbf{H}$ will be denoted by $\|h\|_{\mathbf{H}}$. For any $h \in \mathbf{H}$, we put $W(h) = \int_0^T h(t) dW_t$ and $B(h) = \int_0^T h(t) dB_t$.

For any $m, n \in \mathbb{N}$, denote by $C_p^\infty(\mathbb{R}^{m+n})$ the set of all infinitely differentiable functions $g : \mathbb{R}^{m+n} \to \mathbb{R}$ such that g and all of its partial derivatives have polynomial growth. We make use of the notation $\partial_i g = \frac{\partial g}{\partial x_i}$ whenever $g \in C^1(\mathbb{R}^{m+n})$.

Let $\mathcal{S}$ denote the class of smooth and cylindrical random variables such that a random variable $F \in \mathcal{S}$ has the form

$$F = g(W(h_1), \ldots, W(h_m), B(k_1), \ldots, B(k_n)), \tag{8}$$

where g belongs to $C_p^\infty(\mathbb{R}^{m+n})$, $h_1, \ldots, h_m$ and $k_1, \ldots, k_n$ are in $\mathbf{H}$, and $m, n \in \mathbb{N}$.

For a smooth and cylindrical random variable F of the form (8), its Malliavin derivative with respect to W is the $\mathbf{H}$-valued random variable given by

$$D_t^1 F = \sum_{i=1}^m \partial_i g(W(h_1), \ldots, W(h_m), B(k_1), \ldots, B(k_n)) h_i(t),\ t \in [0, T],$$

and respectively its Malliavin derivative with respect to B is given by

$$D_t^2 F = \sum_{i=1}^n \partial_{m+i} g(W(h_1), \ldots, W(h_m), B(k_1), \ldots, B(k_n)) k_i(t),\ t \in [0, T].$$

For any $p \geq 1$, we will denote the domain of D in $L^p(\Omega)$ by $\mathbb{D}^{1,p}$, meaning that $\mathbb{D}^{1,p}$ is the closure of the class of smooth and cylindrical random variables $\mathcal{S}$ with respect to the norm

$$\|F\|_{1,p} = \left(\mathbb{E}|F|^p + \mathbb{E}\left(\|D^1 F\|_{\mathbf{H}}^2 + \|D^2 F\|_{\mathbf{H}}^2 \right)^{\frac{p}{2}} \right)^{\frac{1}{p}}.$$

We tailor Theorem 2.1.2 in Nualart (2006) to the following lemma which yields a result on the absolute continuity of the law of a random vector with respect to the Lebesgue measure.

Lemma 1. *Let $F = (F_1, F_2)$ be a random vector in $\mathbb{D}^{1,2}$. If the Malliavin matrix $\gamma := (\langle D^1 F_i, D^1 F_j \rangle_{\mathbf{H}} + \langle D^2 F_i, D^2 F_j \rangle_{\mathbf{H}})_{1 \leq i,j \leq 2}$ of F is invertible a.s. Then the law of F is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. Consequently, the joint density of the random variables (F_1, F_2) exists.*

2.3. Bridge Representation for the Joint Density

In this subsection, we show the existence of the joint density of (X_t, Y_t) for any $t > 0$ by using Malliavin calculus. We also give a bridge representation for the joint density by adapting the methodology introduced in Ikeda and Matsumoto Ikeda and Matsumoto (1999).

Theorem 1. *For any $t > 0$, the law of (X_t, Y_t) satisfying (7) is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$. Moreover, the joint probability density $p(t; x, y)$ of (X_t, Y_t) has the following bridge representation*

$$\begin{aligned} & p(t; x, y) \\ = \ & \frac{1}{y\sqrt{2\pi \nu^2 t^{2H}}} e^{-\frac{(\ln(y/y_0))^2}{2\nu^2 t^{2H}}} \times \\ & \frac{1}{2\pi} \int_{\mathbb{R}} \mathbb{E}\left[e^{i\left(x - x_0 - \rho y_0 \int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{\bar{\rho}^2 y_0^2 v_t \xi^2}{2}} \middle| B_t^H = \frac{\ln(y/y_0)}{\nu} \right] d\xi. \end{aligned} \tag{9}$$

where $v_t = \int_0^t e^{2\nu B_s^H} ds$ and $i = \sqrt{-1}$.

Remark 1. *The bridge representation (9) can be regarded as a generalization of the well-known McKean kernel, namely, the classical heat kernel over a 2-dimensional hyperbolic space. For reader's reference, the McKean kernel $p_{\mathbb{H}^2}(t; x, y)$ reads*

$$p_{\mathbb{H}^2}(t;x,y) = \frac{\sqrt{2}e^{-t/8}}{(2\pi t)^{3/2}} \int_d^\infty \frac{\xi e^{-\xi^2/2t}}{\sqrt{\cosh \xi - \cosh d}} d\xi,$$

where $d = d(x, y; x_0, y_0)$ is the geodesic distance from (x, y) to (x_0, y_0). The geodesic distance satisfies $\cosh d(x, y; x_0, y_0) = \frac{(x-x_0)^2+y^2+y_0^2}{2yy_0}$. Note that the McKean kernel is a density with respect to the Riemannian volume form $\frac{1}{y^2}dxdy$. Indeed, in the case where $H = \frac{1}{2}$, $\nu = 1$ and $\rho = 0$, Ikeda-Matsumoto in Ikeda and Matsumoto (1999) showed how to recover the McKean kernel from (9). See also Cheng and Wang (2018) for a different representation in terms of a Bessel bridge for the hyperbolic heat kernel.

Proof. Notice that we can rewrite (7) as

$$\begin{cases} X_t = x_0 + \int_0^t Y_s(\rho dB_s + \bar{\rho} dW_s) - \frac{1}{2}\int_0^t Y_s^2 ds, \\ \\ Y_t = y_0 e^{\nu B_t^H}. \end{cases}$$

Now we fix any $T \geq t$. Then according to Sections 2.2 and 5.2 in Nualart (2006), the Malliavin derivatives of X_t and Y_t are given as follows

$$\begin{aligned} D_\theta^1 Y_t &= 0, \\ D_\theta^2 Y_t &= y_0 \nu e^{\nu B_t^H} K_H(t,\theta) \mathbf{1}_{[0,t]}(\theta) \end{aligned}$$

and

$$\begin{aligned} D_\theta^1 X_t &= \bar{\rho} Y_\theta \mathbf{1}_{[0,t]}(\theta) = \bar{\rho} y_0 e^{\nu B_\theta^H} \mathbf{1}_{[0,t]}(\theta), \\ D_\theta^2 X_t &= \left(\rho Y_\theta + \int_\theta^t \rho D_\theta^2 Y_s dB_s + \int_\theta^t \bar{\rho} D_\theta^2 Y_s dW_s - \int_\theta^t Y_s D_\theta^2 Y_s ds \right) \mathbf{1}_{[0,t]}(\theta) \\ &= \left(\rho y_0 e^{\nu B_\theta^H} + \rho y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dB_s + \bar{\rho} y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dW_s \right) \mathbf{1}_{[0,t]}(\theta) \\ &\quad - y_0^2 \nu \int_\theta^t e^{2\nu B_s^H} K_H(s,\theta) ds \, \mathbf{1}_{[0,t]}(\theta). \end{aligned}$$

Thus, the Malliavin matrix γ of (X_t, Y_t) is given by

$$\gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix},$$

where

$$\begin{aligned} \gamma_{11} &= \int_0^t (D_\theta^1 X_t)^2 d\theta + \int_0^t (D_\theta^2 X_t)^2 d\theta \\ &= \int_0^t \bar{\rho}^2 y_0^2 e^{2\nu B_\theta^H} d\theta + \int_0^t \left(\rho y_0 e^{\nu B_\theta^H} + \rho y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dB_s \right. \\ &\qquad \left. + \bar{\rho} y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dW_s - y_0^2 \nu \int_\theta^t e^{2\nu B_s^H} K_H(s,\theta) ds \right)^2 d\theta, \end{aligned}$$

$$\begin{aligned} \gamma_{12} = \gamma_{21} &= \int_0^t D_\theta^1 X_t D_\theta^1 Y_t d\theta + \int_0^t D_\theta^2 X_t D_\theta^2 Y_t d\theta \\ &= y_0 \nu e^{\nu B_t^H} \int_0^t K_H(t,\theta) \left(\rho y_0 e^{\nu B_\theta^H} + \rho y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dB_s \right. \\ &\qquad \left. + \bar{\rho} y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dW_s - y_0^2 \nu \int_\theta^t e^{2\nu B_s^H} K_H(s,\theta) ds \right) d\theta, \end{aligned}$$

and

$$\begin{aligned} \gamma_{22} &= \int_0^t (D_\theta^1 Y_t)^2 d\theta + \int_0^t (D_\theta^2 Y_t)^2 d\theta \\ &= y_0^2 \nu^2 e^{2\nu B_t^H} \int_0^t K_H(t,\theta)^2 d\theta. \end{aligned}$$

Then it follows from the Cauchy-Schwarz inequality that almost surely

$$\begin{aligned} \gamma_{12}^2 &< y_0^2 \nu^2 e^{2\nu B_t^H} \int_0^t K_H(t,\theta)^2 d\theta \times \\ & \int_0^t \Big(\rho y_0 e^{\nu B_\theta^H} + \rho y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dB_s \\ & \qquad + \bar\rho y_0 \nu \int_\theta^t e^{\nu B_s^H} K_H(s,\theta) dW_s - y_0^2 \nu \int_\theta^t e^{2\nu B_s^H} K_H(s,\theta) ds \Big)^2 d\theta \\ &\le \gamma_{22} \cdot \gamma_{11}, \end{aligned}$$

which implies that the Malliavin matrix γ is invertible a.s. Hence, by Lemma 1 the law of (X_t, Y_t) is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$.

Next, we calculate the joint probability density $p(t; x, y)$ of (X_t, Y_t) as follows. For any bounded and continuous function f defined on $\mathbb{R}^2$, we have

$$\begin{aligned} & \mathbb{E}[f(X_t, Y_t)] \\ = \;& \mathbb{E}\left[f\left(x_0 + y_0 \int_0^t e^{\nu B_s^H} (\rho dB_s + \bar\rho dW_s) - \frac{y_0^2 v_t}{2}, y_0 e^{\nu B_t^H} \right) \right]. \end{aligned} \tag{10}$$

Note that conditioned on $\mathcal{F}_t^B$, $y_0 \bar\rho \int_0^t e^{\nu B_s^H} dW_s$ is normally distributed since W and B^H are independent. Moreover,

$$\begin{aligned} & \mathbb{E}\left[y_0 \bar\rho \int_0^t e^{\nu B_s^H} dW_s \,\middle|\, \mathcal{F}_t^B \right] = 0, \\ & \mathbb{E}\left[\left(y_0 \bar\rho \int_0^t e^{\nu B_s^H} dW_s \right)^2 \,\middle|\, \mathcal{F}_t^B \right] = y_0^2 \bar\rho^2 \int_0^t e^{2\nu B_s^H} ds = y_0^2 \bar\rho^2 v_t. \end{aligned}$$

From (10), it follows by conditioning on $\mathcal{F}_t^B$ that

$$\begin{aligned} & \mathbb{E}[f(X_t, Y_t)] \\ = \;& \mathbb{E}\left[\mathbb{E}\left[f\left(x_0 + y_0 \bar\rho \int_0^t e^{\nu B_s^H} dW_s + y_0 \rho \int_0^t e^{\nu B_s^H} dB_s - \frac{y_0^2 v_t}{2}, y_0 e^{\nu B_t^H} \right) \,\middle|\, \mathcal{F}_t^B \right] \right] \\ = \;& \mathbb{E}\left[\int \left\{ f\left(x_0 + \xi + y_0 \rho \int_0^t e^{\nu B_s^H} dB_s - \frac{y_0^2 v_t}{2}, y_0 e^{\nu B_t^H} \right) \frac{e^{-\frac{\xi^2}{2 y_0^2 \bar\rho^2 v_t}}}{\sqrt{2\pi y_0^2 \bar\rho^2 v_t}} \right\} d\xi \right] \\ = \;& \mathbb{E}\left[\int \left\{ \frac{1}{\sqrt{2\pi y_0^2 \bar\rho^2 v_t}} f\left(x, y_0 e^{\nu B_t^H} \right) e^{-\frac{\left(x - x_0 - y_0 \rho \int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2} \right)^2}{2 y_0^2 \bar\rho^2 v_t}} \right\} dx \right] \\ = \;& \int_{\mathbb{R}^2} f(x,y) \mathbb{E}\left[\frac{1}{\sqrt{2\pi y_0^2 \bar\rho^2 v_t}} e^{-\frac{\left(x - x_0 - y_0 \rho \int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2} \right)^2}{2 y_0^2 \bar\rho^2 v_t}} \,\middle|\, B_t^H = \frac{\ln(y/y_0)}{\nu} \right] \\ & \times \frac{1}{y\sqrt{2\pi \nu^2 t^{2H}}} e^{-\frac{(\ln y - \ln y_0)^2}{2\nu^2 t^{2H}}} dx\, dy. \end{aligned} \tag{11}$$

By using the identity

$$e^{-\frac{u^2}{2y_0^2\bar{\rho}^2 v_t}} = \sqrt{\frac{y_0^2\bar{\rho}^2 v_t}{2\pi}} \int_{\mathbb{R}} e^{iu\xi} e^{-\frac{y_0^2\bar{\rho}^2 v_t \xi^2}{2}} d\xi$$

and letting $u = x - x_0 - \rho y_0 \int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}$, we have

$$\begin{aligned} &\frac{1}{\sqrt{2\pi y_0^2\bar{\rho}^2 v_t}} e^{-\frac{1}{2y_0^2\bar{\rho}^2 v_t}\left(x - x_0 - y_0\rho\int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)^2} \\ = &\frac{1}{2\pi}\int e^{i\left(x - x_0 - \rho y_0 \int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{y_0^2\bar{\rho}^2 v_t \xi^2}{2}} d\xi. \end{aligned} \tag{12}$$

Plugging (12) into the right-hand side of (11), we get

$$\begin{aligned} &\mathbb{E}[f(X_t, Y_t)] \\ = &\frac{1}{y\sqrt{2\pi\nu^2 t^{2H}}} \frac{1}{2\pi} \int_{\mathbb{R}^2} f(x,y) e^{-\frac{(\ln(y/y_0))^2}{2\nu^2 t^{2H}}} \times \\ &\int_{\mathbb{R}} \mathbb{E}\left[e^{i\left(x - x_0 - \rho y_0 \int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{y_0^2\bar{\rho}^2 v_t \xi^2}{2}} \middle| B_t^H = \frac{\ln(y/y_0)}{\nu} \right] d\xi\, dx\, dy. \end{aligned}$$

Finally, we end up with the following bridge representation of the density (9). □

By transforming back to the original variables $(s, a) = (e^x, y)$, we obtain a bridge representation for the joint density $q(t; s, a)$ of (S_t, α_t) in (6).

Corollary 1. *The joint density $q(t; s, a)$ of the lognormal fractional SABR model (6) has the following bridge representation*

$$\begin{aligned} q(t; s, a) &= \frac{e^{-\frac{(\ln(a/a_0))^2}{2\nu^2 t^{2H}}}}{a\sqrt{2\pi\nu^2 t^{2H}}} \frac{1}{2\pi s} \\ &\times \int_{\mathbb{R}} \left(\frac{s}{s_0}\right)^{i\xi} \mathbb{E}\left[e^{i\left(-\rho\int_0^t a_0 e^{\nu B_s^H} dB_s + \frac{a_0^2 v_t}{2}\right)\xi} e^{-\frac{\bar{\rho}^2 a_0^2 v_t}{2}\xi^2} \middle| B_t^H = \frac{\ln(a/a_0)}{\nu} \right] d\xi. \end{aligned} \tag{13}$$

3. Expansion Around Deterministic Path

To gain more intuition and, in particular, a more practical form for applications in obtaining approximations of implied volatility, this section is devoted to deriving an expansion to the lowest order of the bridge representation (9) around a properly chosen deterministic path. The expansion will be shown useful in deriving a small time approximation for implied volatility in Section 4.

Recall that the joint density p of (X_t, Y_t) has the representation given in (9) as

$$\begin{aligned} &p(t; x, y) \\ = &\frac{1}{y\sqrt{2\pi\nu^2 t^{2H}}} e^{-\frac{(\ln(y/y_0))^2}{2\nu^2 t^{2H}}} \times \\ &\frac{1}{2\pi}\int_{\mathbb{R}} \mathbb{E}\left[e^{i\left(x - x_0 - \rho y_0 \int_0^t e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{\bar{\rho}^2 y_0^2 v_t \xi^2}{2}} \middle| B_t^H = \frac{\ln(y/y_0)}{\nu} \right] d\xi. \end{aligned}$$

Let us start with a few naïve calculations, as follows. We expand the above conditional expectation around the deterministic path m_s, for $0 \le s \le t$, that is determined by the conditional expectation of B_s^H given its terminal point $B_t^H = \frac{\ln(y/y_0)}{\nu}$. Precisely,

$$m_s := \mathbb{E}\left[B_s^H \middle| B_t^H = \frac{\ln(y/y_0)}{\nu}\right] = R\left(1, \frac{s}{t}\right)\frac{\ln(y/y_0)}{\nu},$$

where R is defined in (5). By Taylor's expansion, we have, for $n \ge 0$,

$$\begin{aligned} & e^{-i\rho\xi\int_0^t y_0 e^{\nu B_s^H} dB_s} e^{-\frac{1}{2}(\bar\rho^2\xi - i)\xi\int_0^t y_0^2 e^{2\nu B_s^H} ds} \\ \approx\ & e^{-i\rho\xi\int_0^t y_0 e^{\nu m_s} dB_s} e^{-\frac{1}{2}(\bar\rho^2\xi - i)\xi\int_0^t y_0^2 e^{2\nu m_s} ds} \times \\ & \sum_{k,\ell=0}^{n} \frac{(-i\rho\xi)^k}{k!}\left\{\int_0^t y_0\left(e^{\nu B_s^H} - e^{\nu m_s}\right) dB_s\right\}^k \times \\ & \frac{1}{\ell!}\left\{-\frac{1}{2}(\bar\rho^2\xi - i)\xi\int_0^t y_0^2\left(e^{2\nu B_s^H} - e^{2\nu m_s}\right) ds\right\}^\ell. \end{aligned}$$

Thus, even for obtaining a naïve expansion, we shall need a systematic way of computing the conditional expectations of the form, for either $k \ge 1$ or $\ell \ge 1$,

$$\mathbb{E}\left[e^{-i\rho\xi\int_0^t y_0 e^{\nu m_s} dB_s}\left\{\int_0^t\left(e^{\nu B_s^H} - e^{\nu m_s}\right) dB_s\right\}^k\left\{\int_0^t\left(e^{2\nu B_s^H} - e^{2\nu m_s}\right) ds\right\}^\ell \middle| B_t^H = \frac{\ln(y/y_0)}{\nu}\right],$$

which is pretty complicated if not impossible. Nevertheless, as far as the leading order is concerned, small-time expansion of the joint density p to the lowest order (i.e., $k = \ell = 0$) is still manageable. The result is summarized in the following theorem.

In the following sequel, for simplification of the notation, we use $\mathbb{E}_{\frac{\eta}{\nu}}[\cdot]$ to denote $\mathbb{E}[\cdot | B_t^H = \frac{\eta}{\nu}]$, where $\eta = \ln(y/y_0)$. A function g is denoted by $g(t) = O(t^a)$ as $t \to 0^+$ if it satisfies

$$\limsup_{t \to 0^+} \frac{|g(t)|}{t^a} < \infty.$$

Theorem 2. *The joint probability density p of the process (X_t, Y_t) satisfying (7) has the following asymptotic to the lowest order*

$$\begin{aligned} & p(t;x,y) \\ = \ & \frac{1}{2\pi}\frac{1}{y\sqrt{\nu^2 t^{2H}}} e^{-\frac{\eta^2}{2\nu^2 t^{2H}}} \frac{1}{y_0\sqrt{t\psi(\eta)}} e^{-\frac{1}{2y_0^2\psi(\eta)}\left(\frac{x - x_0}{\sqrt{t}} + \frac{y_0^2\sqrt{t}}{2}C_{eR}(\eta) - \rho y_0 t^{-H} C_{RK}(\eta)\frac{\eta}{\nu}\right)^2}\left(1 + O\left(\sqrt{t}\right)\right), \end{aligned} \tag{14}$$

where

$$\begin{aligned} C_{RK}(\eta) &:= \int_0^1 e^{R(1,u)\frac{\eta}{\nu}} K_H(1,u) du, \\ C_{eR}(\eta) &:= \int_0^1 e^{2R(1,u)\frac{\eta}{\nu}} du, \\ \psi(\eta) &:= C_{eR}(\eta) - \rho^2 C_{RK}^2(\eta). \end{aligned}$$

Proof. To the lowest order, p is given by

$$\begin{aligned} & p(t;x,y) \\ = \ & \frac{e^{-\frac{\eta^2}{2\nu^2 t^{2H}}}}{y\sqrt{2\pi\nu^2 t^{2H}}}\frac{1}{2\pi}\int_{\mathbb{R}} e^{i(x - x_0)\xi} e^{-\frac{1}{2}(\bar\rho^2\xi - i)\xi\int_0^t y_0^2 e^{2\nu m_s} ds}\mathbb{E}_{\frac{\eta}{\nu}}\left[e^{-i\rho\xi\int_0^t y_0 e^{\nu m_s} dB_s}\right] d\xi. \end{aligned} \tag{15}$$

We consider the conditional expectation in the above expression. Note that $\int_0^t e^{\nu m_s} dB_s$ and B_t^H are jointly Gaussian. We apply the following identity to evaluate the conditional expectation: if X and Y are jointly normal with mean 0, we can decompose X as

$$X = \frac{\mathrm{cov}(X,Y)}{V(Y)} Y + \sqrt{\frac{V(X)V(Y) - \mathrm{cov}(X,Y)^2}{V(Y)}} Z,$$

where Y and Z are independent and Z is standard normal. Hence,

$$\mathbb{E}[f(X)|Y = y] = \mathbb{E}\left[f\left(\frac{\mathrm{cov}(X,Y)}{V(Y)} y + \sqrt{\frac{V(X)V(Y) - \mathrm{cov}(X,Y)^2}{V(Y)}} Z \right) \right].$$

In our case, $X = \int_0^t e^{\nu m_s} dB_s$ and $Y = B_t^H$, hence

$$\begin{aligned} V(X) &= \int_0^t e^{2\nu m_s} ds = t \int_0^1 e^{2R(1,u)\eta} du = C_{eR}(\eta)\, t, \\ V(Y) &= t^{2H}, \\ \mathrm{cov}(X,Y) &= \int_0^t e^{\nu m_s} K_H(t,s) ds = t^{H+\frac{1}{2}} \int_0^1 e^{R(1,u)\eta} K_H(1,u) du = C_{RK}(\eta)\, t^{H+\frac{1}{2}}. \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbb{E}_{\frac{\eta}{\nu}}\left[e^{-i\rho\xi \int_0^t y_0 e^{\nu m_s} dB_s} \right] \\ = \;& e^{-i\rho\xi y_0 t^{\frac{1}{2}-H} C_{RK}(\eta) \frac{\eta}{\nu}} \mathbb{E}\left[e^{-i\rho\xi y_0 \left\{ \sqrt{t} \sqrt{C_{eR}(\eta) - C_{RK}^2(\eta)} \right\} Z} \right] \\ = \;& \exp\left[-i\rho\xi y_0 t^{\frac{1}{2}-H} C_{RK}(\eta) \frac{\eta}{\nu} - \frac{\rho^2 \xi^2 y_0^2 t}{2} \left\{ C_{eR}(\eta) - C_{RK}^2(\eta) \right\} \right]. \end{aligned}$$

Thus, by substituting the above expression into (15), we obtain

$$\begin{aligned} & p(t;x,y) \\ = \;& \frac{1}{2\pi} \frac{1}{y\sqrt{\nu^2 t^{2H}}} e^{-\frac{\eta^2}{2\nu^2 t^{2H}}} \times \\ & \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i(x-x_0)\xi} e^{-\frac{1}{2}(\bar{\rho}^2\xi - i)\xi t y_0^2 C_{eR}(\eta)} e^{-i\rho\xi y_0 t^{\frac{1}{2}-H} C_{RK}(\eta)\frac{\eta}{\nu} - \frac{\rho^2\xi^2 y_0^2 t}{2}\left\{ C_{eR}(\eta) - C_{RK}^2(\eta) \right\}} d\xi \\ = \;& \frac{1}{2\pi} \frac{1}{y\sqrt{\nu^2 t^{2H}}} e^{-\frac{\eta^2}{2\nu^2 t^{2H}}} \times \\ & \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\left(x - x_0 + \frac{y_0^2 t}{2} C_{eR}(\eta) - \rho y_0 t^{\frac{1}{2}-H} C_{RK}(\eta)\frac{\eta}{\nu} \right)\xi} e^{-\frac{\xi^2 y_0^2 t}{2}\left\{ C_{eR}(\eta) - \rho^2 C_{RK}^2(\eta) \right\}} d\xi \\ = \;& \frac{1}{2\pi} \frac{1}{y\sqrt{\nu^2 t^{2H}}} e^{-\frac{\eta^2}{2\nu^2 t^{2H}}} \frac{1}{y_0\sqrt{t\psi(\eta)}} e^{-\frac{1}{2y_0^2\psi(\eta)}\left(\frac{x-x_0}{\sqrt{t}} + \frac{y_0^2\sqrt{t}}{2} C_{eR}(\eta) - \rho y_0 t^{-H} C_{RK}(\eta)\frac{\eta}{\nu} \right)^2}. \end{aligned} \tag{16}$$

We postpone the detailed error analysis to Appendix A.1 in the Appendix A. □

Remark 2. *We remark that in the logarithmic scale,* (14) *can be expressed in a more concise form as*

$$\begin{aligned} & \ln p(t;x,y) \\ = \;& -\frac{1}{2t^{2H}} \left[\frac{\eta^2}{\nu^2} + \frac{1}{y_0^2 \psi(\eta)} \left(\frac{x - x_0}{t^{\frac{1}{2}-H}} + \frac{y_0^2 t^{H+\frac{1}{2}}}{2} C_{eR}(\eta) - \rho y_0 C_{RK}(\eta) \frac{\eta}{\nu} \right)^2 \right] + O(\ln t). \end{aligned}$$

Remark 3. *In the case that $\nu = 1$, $\rho = 0$, and $H = \frac{1}{2}$, we have*

$$C_{eR}(\eta) = \int_0^1 e^{2R(1,u)\eta} du = \int_0^1 e^{2u\eta} du = \frac{1}{2\eta}(e^{2\eta} - 1) = \frac{y^2 - y_0^2}{2\eta y_0^2}.$$

Then (14) reduces to

$$\begin{aligned}
& \frac{1}{2\pi} \times \frac{1}{y\sqrt{t}} e^{-\frac{\eta^2}{2t}} \times \frac{1}{y_0\sqrt{C_{eR}(\eta)t}} e^{-\frac{1}{2y_0^2 C_{eR}(\eta)t}\left(x - x_0 + \frac{y_0^2 t}{2} C_{eR}(\eta)\right)^2} \\
= \; & \frac{1}{2\pi} \frac{1}{y\sqrt{t}} e^{-\frac{\eta^2}{2t}} \frac{1}{y_0\sqrt{C_{eR}(\eta)t}} e^{-\frac{(x-x_0)^2}{2y_0^2 C_{eR}(\eta)t}} e^{-\frac{x-x_0}{2}} (1 + O(t)) \\
= \; & \frac{1}{2\pi t} e^{-\frac{1}{2t}\left[\eta^2 + \frac{2\eta(x-x_0)^2}{y^2 - y_0^2}\right]} \frac{e^{-\frac{x-x_0}{2}}}{y y_0 \sqrt{C_{eR}(\eta)}} (1 + O(t)).
\end{aligned} \tag{17}$$

Notice that in this case (X_t, Y_t) represents the Brownian motion in the hyperbolic plane whose transition density $p_{\mathbb{H}}$ (with respect to the Riemannian area measure) has the leading term in small time asymptotic as

$$p_{\mathbb{H}}(t; x, y) = \frac{1}{2\pi t} e^{-\frac{d^2(x,y;x_0,y_0)}{2t}} (1 + O(t)),$$

where d denotes the geodesic distance between (x, y) and (x_0, y_0) in the hyperbolic plane. For reader's reference, the hyperbolic cosine of the geodesic distance $d(x, y; x_0, y_0)$ has the closed form expression

$$\cosh d(x, y; x_0, y_0) = \frac{(x - x_0)^2 + y_0^2 + y^2}{2y_0 y}.$$

Thus, in a sense the following function in (17)

$$\tilde{d}(x, y; x_0, y_0) := \sqrt{\eta^2 + \frac{2\eta}{y^2 - y_0^2}(x - x_0)^2}$$

can be regarded as an approximation of the hyperbolic geodesic distance. The complete recovery of the hyperbolic geodesic distance is demonstrated in Section 5 below.

4. Small Time Approximation of Option Price and Implied Volatility

We derive in this section the small-time asymptotics of the premium of a call option and its associated implied volatility by applying the small-time asymptotics for the probability density obtained in Section 3 when $H \le \frac{1}{2}$. It is documented, for example, in Ekström and Lu (2015), that if the underlying asset is governed by an exponential Lévy model, the induced implied volatilities of non-ATM options may explode if jumps exist and the underlying process jumps towards the strike. As we shall see in the following, when $H < \frac{1}{2}$, the small time approximation of implied volatility also explodes; creating a jump-like behavior in the underlying process.

Let $k = \ln K$ be the logmoneyness, t the time to expiry, and recall that $S_t = e^{X_t}$. Though equivalently, we shall be primarily working with the (X_t, Y_t) process as in (7) rather than the (S_t, α_t) process in (6) hereafter. We write the price C of a call as a function of k and t as

$$\begin{aligned}
C(k, t) \; &:= \; \mathbb{E}\left[(S_t - K)^+\right] = \mathbb{E}\left[(e^{X_t} - e^k)^+\right] \\
&= \; \iint (e^x - e^k)^+ p(t; x, y) dx\, dy.
\end{aligned}$$

To evaluate the last integral, we approximate the joint density p by the small time asymptotics obtained in Theorem 2, then, as $t \to 0^+$, apply Laplace asymptotic formula to the resulting integral. For the reader's convenience, we provide proof in Appenidx A.2 a variation of the Laplace asymptotic formula that is tailored for our own use.

Lemma 2. *Let $H \le \frac{1}{2}$. For out-of-money call options, i.e., $k > x_0$, the call price $C(k,t)$ has the following asymptotic as $t \to 0$*

$$\ln C(k,t) \approx -\frac{1}{2t^{2H}} \left\{ \frac{\eta_*^2}{\nu^2} + \frac{1}{y_0^2 \psi(\eta_*)} \left(\frac{k - x_0}{t^{\frac{1}{2}-H}} - \rho y_0 C_{RK}(\eta_*) \frac{\eta_*}{\nu} \right)^2 \right\}, \tag{18}$$

where η_* is the minimizer

$$\eta_* = \operatorname{argmin}\left\{ \eta \in \mathbb{R} : \frac{\eta^2}{\nu^2} + \frac{1}{y_0^2 \psi(\eta)} \left(\frac{k - x_0}{t^{\frac{1}{2}-H}} - \rho y_0 C_{RK}(\eta) \frac{\eta}{\nu} \right)^2 \right\}.$$

Proof. The proof is a straightforward application of the Laplace asymptotic Formula (A12) in Lemma A1. Let $\mathcal{C} = \{(x,\eta) : x \ge k\} \subseteq \mathbb{R}^2$ and $\alpha = \frac{1}{2} - H \ge 0$. By using the asymptotic density (14), consider

$$\begin{aligned}
C(k,t) &= \int_0^\infty \int_k^\infty (e^x - e^k) p(t;x,y) dx dy \\
&= \frac{1}{2\pi} \int_0^\infty \int_k^\infty \left(e^x - e^k\right) \left\{ \frac{1}{y\sqrt{\nu^2 t^{2H}}} e^{-\frac{\eta^2}{2\nu^2 t^{2H}}} \frac{1}{y_0 \sqrt{t\psi(\eta)}} \times \right. \\
&\quad \left. e^{-\frac{1}{2y_0^2 t \psi(\eta)} \left(x - x_0 - \rho y_0 C_{RK}(\eta) \frac{\eta}{\nu} t^{\frac{1}{2}-H} \right)^2} e^{-\frac{x-x_0}{2\psi(\eta)} C_{eR}(\eta)} \left(1 + O\left(\sqrt{t}\right)\right) \right\} dx dy \\
&= \frac{1}{2\pi\nu y_0 t^{H+\frac{1}{2}}} \iint_{\mathcal{C}} \left(\frac{e^x - e^k}{\sqrt{\psi(\eta)}} \right) e^{-\frac{x-x_0}{2\psi(\eta)} C_{eR}(\eta)} \times \\
&\quad e^{-\frac{1}{2t}\left\{ \frac{\eta^2}{\nu^2} t^{2\alpha} + \frac{1}{y_0^2 \psi(\eta)} \left(x - x_0 - \rho y_0 C_{RK}(\eta) \frac{\eta}{\nu} t^\alpha \right)^2 \right\}} \left(1 + O\left(\sqrt{t}\right)\right) dx d\eta.
\end{aligned}$$

Applying the Laplace asymptotic Formula (A12) to the lowest order term in the last expression yields

$$\begin{aligned}
-\ln C(k,t) &\approx \frac{1}{2t} \left\{ \frac{\eta_*^2}{\nu^2} t^{2\alpha} + \frac{1}{y_0^2 \psi(\eta_*)} \left(x_* - x_0 - \rho y_0 C_{RK}(\eta_*) \frac{\eta_*}{\nu} t^\alpha \right)^2 \right\} \\
&= \frac{1}{2t^{2H}} \left\{ \frac{\eta_*^2}{\nu^2} + \frac{1}{y_0^2 \psi(\eta_*)} \left(\frac{x_* - x_0}{t^\alpha} - \rho y_0 C_{RK}(\eta_*) \frac{\eta_*}{\nu} \right)^2 \right\},
\end{aligned}$$

where, for fixed t, (x_*, η_*) is the minimizer of the function

$$\begin{aligned}
(x_*, \eta_*) &= \operatorname{argmin}\left\{ (x,\eta) \in \mathcal{C} : \frac{\eta^2}{\nu^2} t^{2\alpha} + \frac{1}{y_0^2 \psi(\eta)} \left(x - x_0 - \rho y_0 C_{RK}(\eta) \frac{\eta}{\nu} t^\alpha \right)^2 \right\} \\
&= \operatorname{argmin}\left\{ (x,\eta) \in \mathcal{C} : \frac{\eta^2}{\nu^2} + \frac{1}{y_0^2 \psi(\eta)} \left(\frac{x - x_0}{t^\alpha} - \rho y_0 C_{RK}(\eta) \frac{\eta}{\nu} \right)^2 \right\}.
\end{aligned}$$

Since the objective function is continuous in $(x,\eta) \in \mathcal{C}$ and it is a quadratic function in x, it follows that $x_* = k$ when t is small enough, thereby

$$\eta_* = \operatorname{argmin}\left\{ \eta : \frac{\eta^2}{\nu^2} + \frac{1}{y_0^2 \psi(\eta)} \left(\frac{k - x_0}{t^\alpha} - \rho y_0 C_{RK}(\eta) \frac{\eta}{\nu} \right)^2 \right\}.$$

□

Remark 4. *The plots in Figure 1 shows graphically the uniqueness of the minimal point η_* for $H = \frac{1}{4}$ and $H = \frac{3}{4}$. In these particular examples, the contours are convex in the half plane $x > k$, which corresponds to the out-of-money calls. For out-of-money puts, $x < k$, though the contours are not convex, the uniqueness of η_* sustains.*

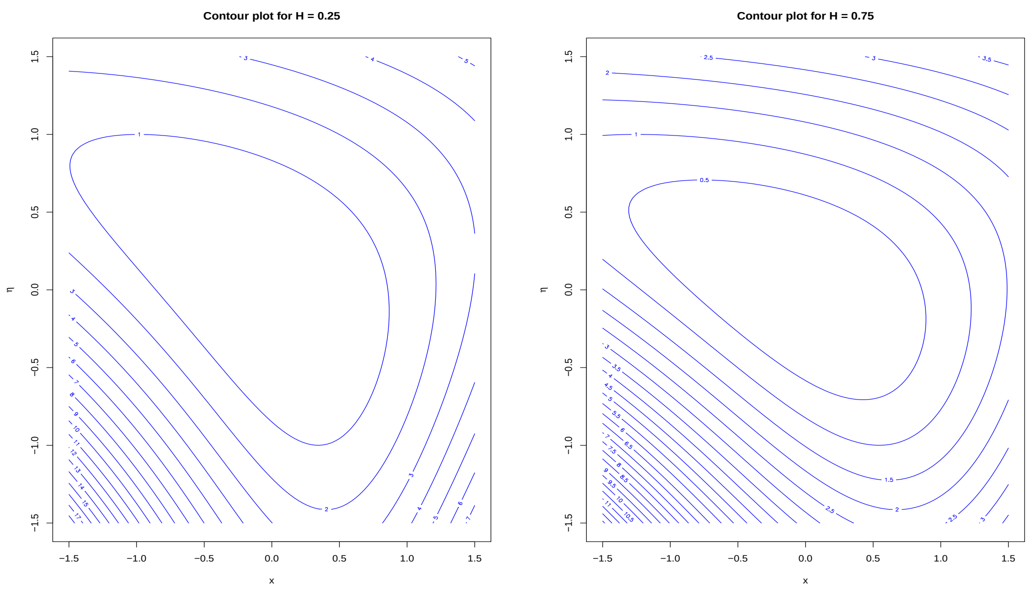

Figure 1. The contour plots. Parameters $\rho = -0.7$, $\nu = 1$, $y_0 = 1$, $t = 0.5$. $H = 0.75$ on the **right**; $H = 0.25$, on the **left**.

So long as we establish an asymptotic for the log price $\ln C(k,t)$ for $k > x_0$, by using the following small time asymptotic for implied volatility in Gao and Lee (2014) or Roper and Rutkowski (2009)

$$\sigma_{\mathrm{BS}}(k,t) = \frac{|k - x_0|}{\sqrt{2t|\ln C(k,t)|}} + O\left(\frac{\ln|\ln C(k,t)|}{\sqrt{t}|\ln C(k,t)|^{3/2}}\right), \tag{19}$$

an asymptotic formula for implied volatility follows immediately. We summarize the result in the following theorem but omitting its proof.

Theorem 3. *Let $H \le \frac{1}{2}$ and let $k = \ln K$ be the log moneyness and $\alpha = \frac{1}{2} - H$. The implied volatility $\sigma_{\mathrm{BS}}(k,t)$ for out-of-money calls ($k > x_0$) has the following asymptotic in small time to expiry*

$$\sigma^2_{\mathrm{BS}}(k,t) = \sigma^2_{\mathrm{BS}}\left(\frac{k}{t^\alpha}\right) \approx \frac{(k-x_0)^2}{t^{2\alpha}}\left\{\frac{\eta_*^2}{\nu^2} + \frac{1}{y_0^2\psi(\eta_*)}\left(\frac{k-x_0}{t^\alpha} - \rho y_0^2 C_{RK}(\eta_*)\frac{\eta_*}{\nu}\right)^2\right\}^{-1}. \tag{20}$$

The minimal point η_* is given Lemma in 2.

Remark 5. *Note that (20) does not recover the SABR formula when $H = \frac{1}{2}$. The derivation of the SABR formula relies heavily on the geometry and symmetry of the underlying SABR plane which is isometric to the Poincaré plane. Figure 2 shows the comparison between the two formulas with time to expiry $t = 1$. Parameters are chosen so as to reproduce the figures in Hagan et al. (2002). In this set of parameters, the maximal difference between the two approximate implied volatility curves is about 1% for logmoneyness $k \in [-1,1]$.*

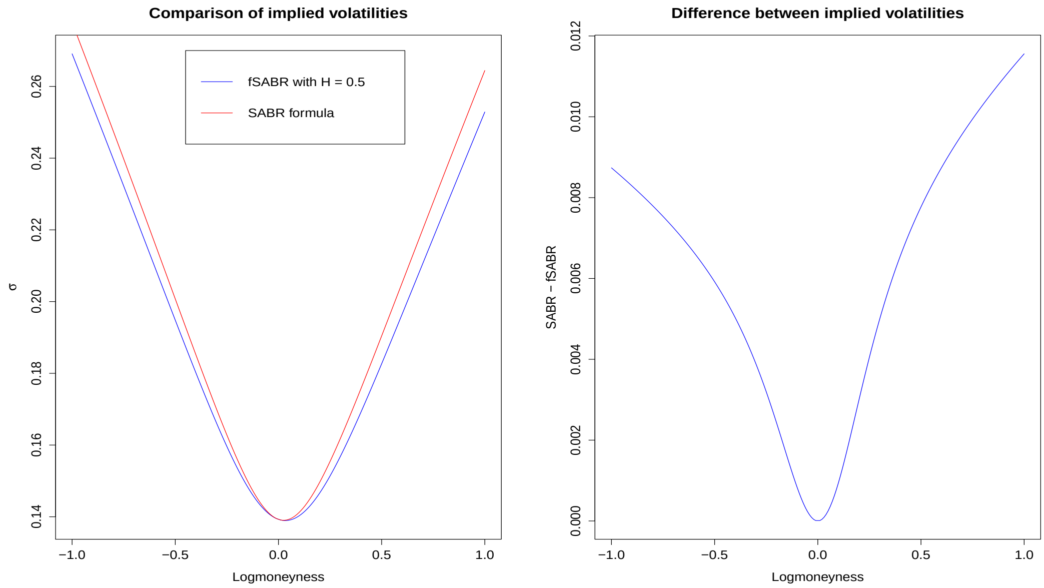

Figure 2. The plot on the **left** shows the approximate implied volatility curves versus logmoneyness with time to expiry $t = 1$ produced by (20) (in blue) the SABR Formula (3) (in red). Parameters are set as $\rho = -0.06867$, $\nu = 0.5778$, $a_0 = 0.13927$. The plot on the **right** shows the difference between the two curves.

We conclude the section by remarking that, as time to expiry t approaches zero, the approximate implied volatility $\sigma_{BS}(k,t)$ flattens out with $H > \frac{1}{2}$; whereas the whole surface $\sigma_{BS}(k,t)$ explodes with $H < \frac{1}{2}$ except for the at-the-money option $k = x_0$. Figure 3 shows the plots of approximate implied volatilities σ given in (20) versus logmoneyness k for time to expiry $t = 0.01$ and $t = 1$ respectively, and various Hurst exponents H. As in Figure 2, parameters are chosen as $a_0 = 0.13927$, $\nu = 0.5778$, and $\rho = -0.06867$. The numerical determination of the η_*'s is relatively efficient since it is basically a one-dimensional optimization problem.

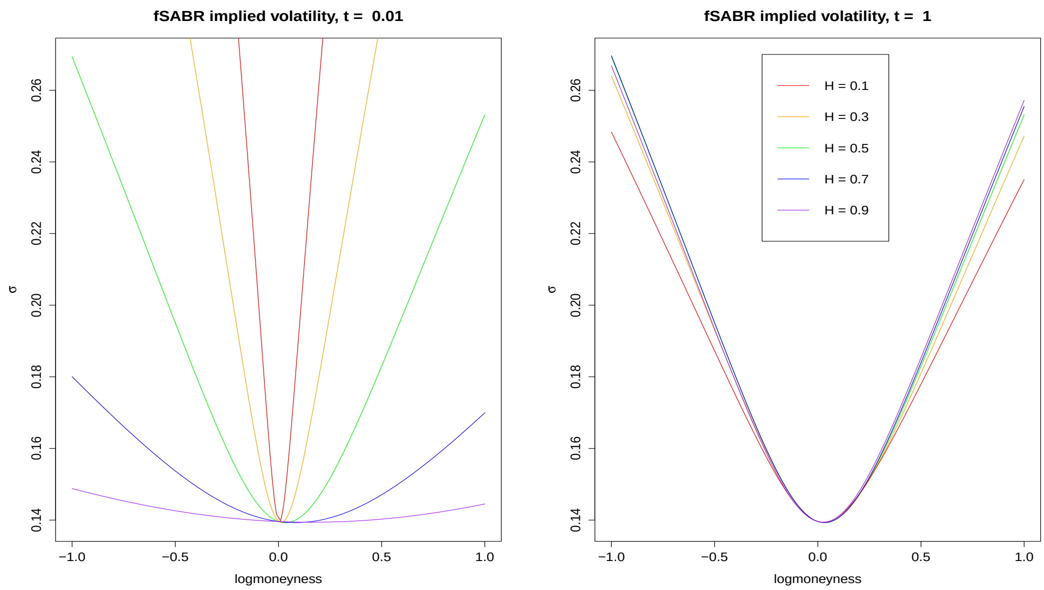

Figure 3. The implied volatility curves for $t = 0.01$ on the **left**, $t = 1$ on the **right**. Parameters are set as $\rho = -0.06867$, $\nu = 0.5778$, $a_0 = 0.13927$. $H = 0.1$ in red, $H = 0.3$ in orange, $H = \frac{1}{2}$ in green, $H = 0.7$ in blue, $H = 0.9$ in purple.

5. A Heuristic Large Deviation Principle

In this section, we provide a heuristic derivation of the sample path large deviation principle for (X_t, Y_t) in small time by bootstrapping the bridge representation to multi-period. For simplicity, we introduce the following vector notations.

$$
\begin{aligned}
&\boldsymbol{t} = (t_1, \ldots, t_n), \qquad \boldsymbol{x_t} = (x_{t_1}, \ldots, x_{t_n}), \qquad \boldsymbol{y_t} = (y_{t_1}, \ldots, y_{t_n}),\\
&\boldsymbol{B_t^H} = (B_{t_1}^H, \ldots, B_{t_n}^H), \qquad \boldsymbol{X_t} = (X_{t_1}, \ldots, X_{t_n}), \qquad \boldsymbol{Y_t} = (Y_{t_1}, \ldots, Y_{t_n}),\\
&\boldsymbol{\xi_t} = (\xi_{t_1}, \ldots, \xi_{t_n}), \qquad \boldsymbol{\eta_t} = (\eta_{t_1}, \ldots, \eta_{t_n}), \qquad \boldsymbol{\zeta_t} = (\zeta_{t_1}, \ldots, \zeta_{t_n}).
\end{aligned}
$$

Theorem 4. *The multiperiod joint density p of (X_t, Y_t)*

$$
p(x_1, y_1, \ldots, x_n, y_n) := \mathbb{P}[(X_{t_1}, Y_{t_1}) = (x_1, y_1), \ldots, (X_{t_n}, Y_{t_n}) = (x_n, y_n)]
$$

has the following bridge representation

$$
\begin{aligned}
&p(x_1, y_1, \ldots, x_n, y_n) \\
= \;& \mathbb{E}\left[\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi y_0^2 \bar{\rho}^2 \Delta v_{t_k}}} e^{-\frac{1}{2y_0^2\bar{\rho}^2\Delta v_{t_k}}\left(\Delta x_{t_k} - y_0\rho\int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s + \frac{y_0^2}{2}\Delta v_{t_k}\right)^2} \middle| \nu \boldsymbol{B_t^H} = \boldsymbol{\eta_t}\right] \times \\
& \mathbb{P}\left[y_0 e^{\nu \boldsymbol{B_t^H}} = \boldsymbol{y_t}\right],
\end{aligned}
\tag{21}
$$

where $\boldsymbol{\eta_t} = \log \boldsymbol{y_t} - \log y_0$, $\Delta x_{t_k} = x_{t_k} - x_{t_{k-1}}$, and $\Delta v_{t_k} = v_{t_k} - v_{t_{k-1}}$ for $k = 1, \ldots, n$. Recall that $v_t = \int_0^t e^{\nu B_s^H} ds$.

Proof. For any bounded measurable function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, consider the expectation

$$
\begin{aligned}
& \iint f(\boldsymbol{x_t}, \boldsymbol{y_t}) p(\boldsymbol{x_t}, \boldsymbol{y_t}) d\boldsymbol{x_t} d\boldsymbol{y_t} \\
=\; & \mathbb{E}[f(\boldsymbol{X_t}, \boldsymbol{Y_t})] \\
=\; & \mathbb{E}\left[\mathbb{E}\left[f(\boldsymbol{X_t}, \boldsymbol{Y_t}) | \mathcal{F}_{t_n}^B\right]\right].
\end{aligned}
$$

Let $\xi_{t_i} = \int_0^{t_i} e^{\nu B_s^H} dW_s$, $\zeta_{t_i} = \int_0^{t_i} e^{\nu B_s^H} dB_s$ and thus accordingly $\Delta\xi_{t_i} = \xi_{t_i} - \xi_{t_{i-1}} = \int_{t_{i-1}}^{t_i} e^{\nu B_s^H} dW_s$, $\Delta\zeta_{t_i} = \zeta_{t_i} - \zeta_{t_{i-1}} = \int_{t_{i-1}}^{t_i} e^{\nu B_s^H} dB_s$. Note that, conditioned on $\mathcal{F}_{t_n}^B$, the random variables $\Delta\xi_{t_i}$'s are independent normal with mean 0 and variance Δv_{t_i}. We calculate the conditional expectation as follows.

$$
\begin{aligned}
& \mathbb{E}\left[f(\boldsymbol{X_t}, \boldsymbol{Y_t}) | \mathcal{F}_{t_n}^B\right] \\
=\; & \mathbb{E}\left[f\left(x_0 + \rho y_0 \boldsymbol{\zeta_t} - \frac{y_0^2}{2}\boldsymbol{v_t} + \bar{\rho} y_0 \boldsymbol{\xi_t}, y_0 e^{\nu \boldsymbol{B_t^H}}\right) \middle| \mathcal{F}_{t_n}^B\right] \\
=\; & \int f\left(x_0 + \rho y_0 \boldsymbol{\zeta_t} - \frac{y_0^2}{2}\boldsymbol{v_t} + \bar{\rho} y_0 \boldsymbol{\xi_t}, y_0 e^{\nu \boldsymbol{B_t^H}}\right) \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\Delta v_{t_k}}} e^{-\frac{(\Delta\xi_{t_k})^2}{2\Delta v_{t_k}}} d\Delta\boldsymbol{\xi_t}.
\end{aligned}
\tag{22}
$$

By applying the change of variables

$$
x_{t_k} = x_0 + \rho y_0 \zeta_{t_k} - \frac{y_0^2}{2} v_{t_k} + \bar{\rho} y_0 \xi_{t_k},
$$

thus

$$
\Delta\xi_{t_k} = \frac{1}{\bar{\rho} y_0}\left(\Delta x_{t_k} - \Delta\zeta_{t_k} - \frac{y_0^2}{2}\Delta v_{t_k}\right).
$$

The integral (22) becomes

$$
\begin{aligned}
& \int f\left(\boldsymbol{x_t}, y_0 e^{\nu \boldsymbol{B}_t^H}\right) \prod_{k=1}^n \frac{1}{\sqrt{2\pi \Delta v_{t_k}}} e^{-\frac{(\Delta \xi_{t_k})^2}{2\Delta v_{t_k}}} \, d\Delta \boldsymbol{\xi_t} \\
= & \int f\left(\boldsymbol{x_t}, y_0 e^{\nu \boldsymbol{B}_t^H}\right) \prod_{k=1}^n \frac{1}{\sqrt{2\pi \bar{\rho}^2 y_0^2 \Delta v_{t_k}}} e^{-\frac{1}{2\bar{\rho} y_0 \Delta v_{t_k}}\left(\Delta x_{t_k} - \rho y_0 \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s - \frac{y_0^2}{2}\Delta v_{t_k}\right)^2} d\Delta \boldsymbol{x_t} \\
= & \int f\left(\boldsymbol{x_t}, y_0 e^{\nu \boldsymbol{B}_t^H}\right) \prod_{k=1}^n \frac{1}{\sqrt{2\pi \bar{\rho}^2 y_0^2 \Delta v_{t_k}}} e^{-\frac{1}{2\bar{\rho} y_0 \Delta v_{t_k}}\left(\Delta x_{t_k} - \rho y_0 \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s - \frac{y_0^2}{2}\Delta v_{t_k}\right)^2} d\boldsymbol{x_t}
\end{aligned}
$$

since the Jacobian between $d\Delta \boldsymbol{x_t}$ and $d\boldsymbol{x_t}$ 1. It follows that

$$
\begin{aligned}
& \iint f(\boldsymbol{x_t}, \boldsymbol{y_t}) p(\boldsymbol{x_t}, \boldsymbol{y_t}) d\boldsymbol{x_t} d\boldsymbol{y_t} \\
= & \mathbb{E}\left[\mathbb{E}[f(\boldsymbol{X_t}, \boldsymbol{Y_t}) | \mathcal{F}_t^B]\right] \\
= & \mathbb{E}\left[\int f\left(\boldsymbol{x_t}, y_0 e^{\nu \boldsymbol{B}_t^H}\right) \prod_{k=1}^n \frac{1}{\sqrt{2\pi \bar{\rho}^2 y_0^2 \Delta v_{t_k}}} e^{-\frac{1}{2\bar{\rho} y_0 \Delta v_{t_k}}\left(\Delta x_{t_k} - \rho y_0 \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s - \frac{y_0^2}{2}\Delta v_{t_k}\right)^2} d\boldsymbol{x_t}\right] \\
= & \iint d\boldsymbol{x_t} d\boldsymbol{y_t} f(\boldsymbol{x_t}, \boldsymbol{y_t}) \times \\
& \mathbb{E}\left[\prod_{k=1}^n \frac{1}{\sqrt{2\pi \bar{\rho}^2 y_0^2 \Delta v_{t_k}}} e^{-\frac{1}{2\bar{\rho} y_0 \Delta v_{t_k}}\left(\Delta x_{t_k} - \rho y_0 \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s - \frac{y_0^2}{2}\Delta v_{t_k}\right)^2} \middle| \nu \boldsymbol{B}_t^H = \boldsymbol{\eta_t}\right] \times \\
& \mathbb{P}\left[y_0 e^{\nu \boldsymbol{B}_t^H} = \boldsymbol{y_t}\right].
\end{aligned}
$$

This completes the proof of bridge representation (21) since f is arbitrary. □

To move onto a heuristic derivation of the sample path large deviation principle for (X_t, Y_t) in small time, we take logarithm on both sides of (21) and obtain

$$
\begin{aligned}
& \log p(x_{t_1}, y_{t_1}, \ldots, x_{t_n}, y_{t_n}) \\
= & \log \mathbb{E}\left[\prod_{k=1}^n \frac{1}{\sqrt{2\pi y_0^2 \bar{\rho}^2 \Delta v_{t_k}}} e^{-\frac{1}{2y_0^2 \bar{\rho}^2 \Delta v_{t_k}}\left(\Delta x_{t_k} - y_0 \rho \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s + \frac{y_0^2}{2}\Delta v_{t_k}\right)^2} \middle| \nu \boldsymbol{B}_t^H = \boldsymbol{\eta}\right] \\
& + \log \mathbb{P}\left[\nu \boldsymbol{B}_t^H = \boldsymbol{\eta_t}\right] - \sum \log y_{t_i}. \qquad (23)
\end{aligned}
$$

In the following, we ignore the last term on the right-hand side of (23) and intuitively calculate the limits as $n \to \infty$ of the first two terms. Note that to the leading order we have

$$
\log \mathbb{P}\left[\nu \boldsymbol{B}_t^H = \boldsymbol{\eta_t}\right] \approx -\frac{1}{2\nu^2} \boldsymbol{\eta}' \boldsymbol{R}^{-1} \boldsymbol{\eta},
$$

where $\boldsymbol{R} = [R(t_i, t_j)]$ is the covariance matrix of $\boldsymbol{B}_t^H$. We further discretize the autovariance R of fractional Brownian motion as

$$
\begin{aligned}
R(t_i, t_j) = \mathbb{E}\left[B_{t_i}^H B_{t_i}^H\right] &= \int_0^{t_i \wedge t_j} K_H(t_i, s) K_H(t_j, s) ds \\
\approx \sum_{k=0}^{i \wedge j} K_H(t_i, t_k) K_H(t_j, t_k) \Delta t &= \boldsymbol{K}' \boldsymbol{K} \Delta t,
\end{aligned}
$$

where $\boldsymbol{K}$ denotes the upper triangular matrix

$$\boldsymbol{K}_{ij} = \begin{cases} K_H(t_i, t_j), & \text{if } i \geq j; \\ 0, & \text{otherwise.} \end{cases}$$

Thereby, $\boldsymbol{R}^{-1} = \frac{1}{\Delta t}\boldsymbol{K}^{-1}(\boldsymbol{K}')^{-1}$. Let $\boldsymbol{b} = (b_{t_1}, \ldots b_{t_n})$ be the solution to the linear system

$$\frac{\boldsymbol{\eta}}{\nu} = \boldsymbol{K}\boldsymbol{b}\Delta t.$$

It follows that

$$\begin{aligned} & \frac{1}{2\nu^2}\boldsymbol{\eta}'\boldsymbol{R}^{-1}\boldsymbol{\eta} = \frac{1}{2}\Delta t \boldsymbol{b}'\boldsymbol{K}'\boldsymbol{R}^{-1}\boldsymbol{K}\boldsymbol{b}\Delta t \\ = & \frac{1}{2}\boldsymbol{b}'\boldsymbol{b}\Delta t = \frac{1}{2}\sum_{k=1}^{n} b_{t_k}^2 \Delta t \\ \longrightarrow & \frac{1}{2}\int_0^T b_t^2 dt \quad \text{as } n \to \infty. \end{aligned}$$

Also in the limit as $n \to \infty$, we obtain $\eta_t = \nu \int_0^t K_H(t,s) b_s ds$.

On the other hand, for the first term on the right-hand side of (23), we have

$$\begin{aligned} & \log \mathbb{E}\left[\prod_{k=1}^{n} \frac{1}{\sqrt{2\pi y_0^2 \bar{\rho}^2 \Delta v_{t_k}}} e^{-\frac{1}{2y_0^2\bar{\rho}^2\Delta v_{t_k}}\left(\Delta x_{t_k} - y_0\rho\int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s + \frac{y_0^2}{2}\Delta v_{t_k}\right)^2} \middle| \nu \boldsymbol{B}^H = \boldsymbol{\eta} \right] \\ \approx & \sum_{k=1}^{n} \mathbb{E}\left[-\frac{1}{2y_0^2\bar{\rho}^2\Delta v_{t_k}} \left(\Delta x_{t_k} - y_0\rho \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s \right)^2 \middle| \nu \boldsymbol{B}^H = \boldsymbol{\eta} \right]. \end{aligned}$$

Note that conditioned on $\nu \boldsymbol{B}^H = \boldsymbol{\eta}$, we have

$$\Delta v_{t_k} = \int_{t_{k-1}}^{t_k} e^{2\nu B_s^H} ds \approx e^{2\eta_{t_{k-1}}} \Delta t = e^{2\nu \sum_{j=0}^{k-1} K_H(t_{k-1}, t_j) b_{t_j} \Delta t} \Delta t$$

as well as

$$\begin{aligned} & \Delta x_{t_k} - y_0\rho \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s \approx \Delta x_{t_k} - y_0\rho e^{\eta_{t_{k-1}}} b_{t_{k-1}} \Delta t \\ = & \left(\frac{\Delta x_{t_k}}{\Delta t} - y_0 \rho e^{\nu \sum_{j=0}^{k-1} K_H(t_{k-1}, t_j) b_{t_j} \Delta t} b_{t_{k-1}} \right) \Delta t. \end{aligned}$$

It follows that the first term in (23) has the limit

$$\begin{aligned} & \sum_{k=1}^{n} \mathbb{E}\left[-\frac{1}{2y_0^2\bar{\rho}^2\Delta v_{t_k}} \left(\Delta x_{t_k} - y_0\rho \int_{t_{k-1}}^{t_k} e^{\nu B_s^H} dB_s \right)^2 \middle| \nu \boldsymbol{B}^H = \boldsymbol{\eta} \right] \\ \approx & -\sum_{k=0}^{n} \frac{1}{2y_0^2 \bar{\rho}^2 e^{2\nu \sum_{j=0}^{k-1} K_H(t_{k-1}, t_j) b_{t_j} \Delta t}} \left(\frac{\Delta x_{t_k}}{\Delta t} - y_0 \rho e^{\nu \sum_{j=0}^{k-1} K_H(t_{k-1}, t_j) b_{t_j} \Delta t} b_{t_{k-1}} \right)^2 \Delta t \\ \longrightarrow & -\frac{1}{2} \int_0^T \frac{1}{y_0^2 \bar{\rho}^2 e^{2\nu \int_0^t K_H(t,s) b_s ds}} \left(\dot{x}_t - y_0 \rho e^{\nu \int_0^t K_H(t,s) b_s ds} b_t \right)^2 dt \end{aligned}$$

as $n \to \infty$.

Putting the two limits together, we obtain heuristically for $T \approx 0$ that

$$\begin{aligned}
& -\log \mathbb{P}[X_t = x_t, Y_t = y_t, \text{ for } t \in [0,T]] \\
\approx\ & \frac{1}{2}\int_0^T \frac{1}{y_0^2 \bar{\rho}^2 e^{2\nu \int_0^t K_H(t,s) b_s ds}} \left(\dot{x}_t - y_0 \rho e^{\nu \int_0^t K_H(t,s) b_s ds} b_t \right)^2 dt + \frac{1}{2}\int_0^T b_t^2 dt \\
=\ & \frac{1}{2}\int_0^T \frac{1}{\bar{\rho}^2 y_t^2} (\dot{x}_t - \rho y_t b_t)^2 dt + \frac{1}{2}\int_0^T b_t^2 dt \\
=\ & \frac{1}{2}\int_0^T \frac{1}{\bar{\rho}^2} \left(\frac{\dot{x}_t}{y_t} - \rho b_t \right)^2 dt + \frac{1}{2}\int_0^T b_t^2 dt, \qquad (24)
\end{aligned}$$

where $b \in L^2[0,T]$ satisfies the integral equation

$$\log y_t - \log y_0 = \nu \int_0^t K_H(t,s) b_s ds$$

for all $t \in [0,T]$. We remark that (24) should serve as the rate function for the sample path large deviation principle in small time for (X_t, Y_t). Moreover, one may define the "geodesic" from the initial point (x_0, y_0) to the terminal point (x_T, y_T) in the fSABR plane as the path (x_t^*, y_t^*) which minimizes the functional (24), i.e.,

$$(x_t^*, y_t^*) := \underset{t \mapsto (x_t, y_t)}{\operatorname{argmin}} \frac{1}{2}\int_0^T \frac{1}{\bar{\rho}^2} \left(\frac{\dot{x}_t}{y_t} - \rho b_t \right)^2 dt + \frac{1}{2}\int_0^T b_t^2 dt,$$

where again b_t is determined by solving the integral equation

$$\log y_t - \log y_0 = \nu \int_0^t K_H(t,s) b_s ds. \qquad (25)$$

Also, the minimizer can be regarded as the "geodesic" connecting (x_0, y_0) and (x_T, y_T).

Remark 6. *Note that b_t is indeed determined by the inverse operator K_H^{-1} applied to $\log \frac{y_t}{y_0}$. In particular, with $H = \frac{1}{2}$ this inverse operator reduces to the usual derivative. Thus, with $H = \frac{1}{2}$,*

$$b_t = \frac{d}{dt}\left(\log \frac{y_t}{y_0} \right) = \frac{\dot{y}_t}{y_t}.$$

The functional (24) becomes

$$\begin{aligned}
& \frac{1}{2}\int_0^T \frac{1}{\bar{\rho}^2} \left(\frac{\dot{x}_t}{y_t} - \rho b_t \right)^2 dt + \frac{1}{2}\int_0^T b_t^2 dt \\
=\ & \frac{1}{2}\int_0^T \frac{1}{\bar{\rho}^2} \left(\frac{\dot{x}_t}{y_t} - \rho \frac{\dot{y}_t}{y_t} \right)^2 dt + \frac{1}{2}\int_0^T \frac{\dot{y}_t^2}{y_t^2} dt \\
=\ & \frac{1}{2}\int_0^T \frac{1}{\bar{\rho}^2 y_t^2} \left(\dot{x}_t^2 - 2\rho \dot{x}_t \dot{y}_t + \dot{y}_t^2 \right) dt.
\end{aligned}$$

The last expression is the energy functional (up to the constant factor $\frac{1}{2}$) associated with the Riemann metric $ds^2 = \frac{1}{\bar{\rho}^2 y^2}(dx^2 - 2\rho dx dy + dy^2)$. The diffusion process associated with this Riemann metric is governed by the SDEs

$$\begin{aligned}
dX_t &= Y_t dW_t, \\
dY_t &= Y_t dZ_t,
\end{aligned}$$

where W_t and Z_t are correlated Brownian motion with constant correlation ρ, which up to a linear transformation is the upper plane model of the Poincaré space. In other words, with $H = \frac{1}{2}$, the functional (24) recovers the energy functional for the classical Poincaré space, which is isometric to the SABR plane.

Lastly, with the aid of the sample path large deviation principle (24), it is nearly a common practice, say by applying the Laplace asymptotic formula, to conclude that the log premium of an out-of-money call in small time has the asymptotic as $t \to 0$

$$-\log C(k,t) \approx -\log \mathbb{P}[X_t \geq k] \approx \frac{1}{2}\int_0^T \frac{1}{\bar{\rho}^2}\left(\frac{\dot{x}_t^*}{y_t^*} - \rho b_t^*\right)^2 dt + \frac{1}{2}\int_0^T b_t^{*2} dt,$$

where (x_t^*, y_t^*, b_t^*) denotes the optimal path that minimizes the functional (24) subject to the constraint $x_t^* = k$ and y_t^*, b_t^* satisfy the integral Equation (25). Thus, by applying (19), an approximation of implied volatility in small time is readily obtained. We summarize the result in the following proposition which, with $H = \frac{1}{2}$, recovers the SABR Formula (3). However, for $H \neq \frac{1}{2}$, the numerical implementation of (26) is more involved than that of (20) since, as opposed to a one dimensional optimization problem, it is subject to solving a two-dimensional constrained variational problem.

Proposition 1 (fSABR formula)**.** *Let $k = \log\left(\frac{K}{s_0}\right)$ be the log moneyness. The implied volatility $\sigma_{\mathrm{BS}}(k,t)$ in a small time to expiry has the asymptotics*

$$\sigma_{\mathrm{BS}}^2 \approx \frac{k^2}{T}\left(\int_0^T \left\{\frac{1}{\bar{\rho}^2 y_t^{*2}}(\dot{x}_t^* - \rho y_t^* b_t^*)^2 + b_t^{*2}\right\} dt\right)^{-1}, \tag{26}$$

where (x_t^*, b^*) is the minimizer of the variational problem

$$(x^*, b^*) = \mathit{argmin}\left\{\dot{x}, b \in L^2[0,T] : \int_0^T \left(\frac{1}{\bar{\rho}^2 y_t^2}(\dot{x}_t - \rho y_t b_t)^2 + b_t^2\right) dt\right\}$$

with $x_T = k$ and y_t^* satisfying

$$\log y_t^* - \log y_0 = \nu \int_0^t K_H(t,s) b_s^* ds$$

for $t \in [0,T]$. Notice that (26) recovers the SABR Formula (3) with $H = \frac{1}{2}$.

6. Conclusions and Discussion

We showed in this paper a bridge representation in Fourier space and a small time asymptotic for the joint probability of lognormal fractional SABR model for general $\rho \in (-1,1)$. An application of the asymptotics of the joint density is an approximation of the implied volatility in a short time. Due to the different nature of methodologies, the newly obtained approximation of implied volatilities in small time does not recover the celebrated SABR formula for implied volatility (to the zeroth order) when the Hurst exponent H equals a half. To recover the SABR formula, we presented a heuristic derivation of the sample path large deviation principle for the lognormal fractional SABR model by bootstrapping via the multiperiod joint density. We emphasize once again that the same trick is applicable to general fractional SABR models, i.e., to include a local volatility component in the process S_t for an underlying asset. We leave the rigorous proof of the sample path large deviation principle for fractional SABR models in future work. Lastly, the bridge representation methodology is also applicable to the case in which the volatility process is governed by an exponential fractional Ornstein-Uhlenbeck process since a fractional Ornstein-Uhlenbeck process is Gaussian as well. However, as the time to expiry approaches zero, the mean reversion part does not really play a role in the large deviation regime.

Author Contributions: All the authors contributed equally to the article in every aspect. All authors have read and agreed to the published version of the manuscript.

Funding: Jiro Akahori is supported by JSPS KAKENHI Grant Number 23330109, 24340022, 23654056, 25285102, 20K03666, and the project RARE-318984 (an FP7 Marie Curie IRSES). Tai-Ho Wang is partially supported by the PSC-CUNY52 Award and the National Natural Science Foundation of China grant 11971040.

Acknowledgments: We are grateful for helpful discussions with the participants of the conferences: At the Frontiers of Quantitative Finance at International Centre of Mathematical Sciences, Edinburgh, UK, and Mathematics of Quantitative Finance at Mathematisches Forschungsinstitut Oberwolfach, Oberwolfach, Germany. We would also like to thank the anonymous referees for their comments and suggestions which helped improve the readability and presentation of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Technical Proofs

In the appendix, we provide a detailed error analysis of the asymptotic expansion for (14) and a version of Laplace's asymptotic formula that is readily applicable to our case.

Appendix A.1. Error Analysis

Let $C_0^\infty(\mathbb{R}^2)$ be the space of smooth functions defined on $\mathbb{R}^2$ with compact support. For a given $f \in C_0^\infty(\mathbb{R}^2)$, recalling $\eta = \ln(y/y_0)$, from (9) we have

$$
\begin{aligned}
&\mathbb{E}[f(X_t, Y_t)] = \iint f(x,y) p(t;x,y) dx\, dy \\
=\;& \frac{1}{2\pi} \iiint f(x,y) \frac{e^{-\frac{\eta^2}{2\nu^2 t^{2H}}}}{y\sqrt{2\pi\nu^2 t^{2H}}} e^{i(x-x_0)\xi} \mathbb{E}_{\frac{\eta}{\nu}}\left[e^{i\left(-\rho\int_0^t y_0 e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{\bar\rho^2 y_0^2 v_t \xi^2}{2}} \right] d\xi dx dy \\
=\;& \frac{1}{2\pi} \iint e^{-ix_0\xi} \hat f(\xi, y) \frac{e^{-\frac{\eta^2}{2\nu^2 t^{2H}}}}{y\sqrt{2\pi\nu^2 t^{2H}}} \mathbb{E}_{\frac{\eta}{\nu}}\left[e^{i\left(-\rho\int_0^t y_0 e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{\bar\rho^2 y_0^2 v_t \xi^2}{2}} \right] dy d\xi \\
=\;& \frac{1}{2\pi} \int e^{-ix_0\xi} \mathbb{E}\left[\hat f(\xi, Y_t) e^{i\left(-\rho\int_0^t y_0 e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{\bar\rho^2 y_0^2 v_t \xi^2}{2}} \right] d\xi,
\end{aligned} \tag{A1}
$$

where

$$
\hat f(\xi, y) = \int e^{i\xi x} f(x,y) dx
$$

is the Fourier transform of f with respect to x.

Note that the right-hand side of (14) equals the right-hand side of (15). We compare (A1) with the following expression obtained by using the approximate joint density in (14) and obtain

$$
\begin{aligned}
& \frac{1}{2\pi} \iiint f(x,y) \frac{e^{-\frac{\eta^2}{2\nu^2 t^{2H}}}}{y\sqrt{2\pi\nu^2 t^{2H}}} \\
&\qquad \times e^{i(x-x_0)\xi} e^{-\frac{1}{2}(\bar\rho^2\xi - i)\xi \int_0^t y_0^2 e^{2\nu m_s} ds} \mathbb{E}_{\frac{\eta}{\nu}}\left[e^{-i\rho\xi\int_0^t y_0 e^{\nu m_s} dB_s} \right] d\xi dx dy \\
=\;& \frac{1}{2\pi} \iint e^{-ix_0\xi} \hat f(\xi, y) \frac{e^{-\frac{\eta^2}{2\nu^2 t^{2H}}}}{y\sqrt{2\pi\nu^2 t^{2H}}} \\
&\qquad \times \mathbb{E}_{\frac{\eta}{\nu}}\left[e^{i\left(-\rho\int_0^t y_0 e^{\nu m_s} dB_s + \frac{y_0^2}{2}\int_0^t e^{2\nu m_s} ds\right)\xi} e^{-\frac{\bar\rho^2 y_0^2 \xi^2}{2}\int_0^t e^{2\nu m_s} ds} \right] d\xi dx dy \\
=\;& \frac{1}{2\pi} \int e^{-ix_0\xi} \mathbb{E}\left[\hat f(\xi, Y_t) e^{i\left(-\rho\int_0^t y_0 e^{\nu m_s} dB_s + \frac{y_0^2}{2}\int_0^t e^{2\nu m_s} ds\right)\xi} e^{-\frac{\bar\rho^2 y_0^2 \xi^2}{2}\int_0^t e^{2\nu m_s} ds} \right] d\xi.
\end{aligned} \tag{A2}
$$

For simplification, denote

$$\lambda_1(t) = e^{i\left(-\rho\int_0^t y_0 e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi} e^{-\frac{\bar{\rho}^2 y_0^2 v_t}{2}\xi^2}$$

and

$$\lambda_2(t) = e^{i\left(-\rho\int_0^t y_0 e^{\nu m_s} dB_s + \frac{y_0^2}{2}\int_0^t e^{2\nu m_s} ds\right)\xi} e^{-\frac{\bar{\rho}^2 y_0^2 \xi^2}{2}\int_0^t e^{2\nu m_s} ds}.$$

Then the modulus of the difference between (A1) and (A2) is equal to

$$\left|\frac{1}{2\pi}\int e^{-ix_0\xi}\mathbb{E}\left[\hat{f}(\xi, Y_t)(\lambda_1(t) - \lambda_2(t))\right]d\xi\right|. \tag{A3}$$

The goal is to show that (A3) converges to zero in the order of $t^{\frac{1}{2}}$ as $t \to 0$, for every $f \in C_0^\infty(\mathbb{R}^2)$.

By applying the following inequality, for any $z, w \in \mathbb{C}$,

$$|e^z - e^w| \le \left(e^{\Re(z)} + e^{\Re(w)}\right)|z - w|,$$

where $\Re(z)$ denotes the real part of z, we have

$$\begin{aligned}
&|\lambda_1(t) - \lambda_2(t)| \\
\le\ & \left(e^{-\frac{\bar{\rho}^2 y_0^2 v_t \xi^2}{2}} + e^{-\frac{\bar{\rho}^2 y_0^2 \xi^2}{2}\int_0^t e^{2\nu m_s} ds}\right) \times \\
& \left| i\left(-\rho\int_0^t y_0 e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2}\right)\xi - \frac{\bar{\rho}^2 y_0^2 v_t \xi^2}{2} \right. \\
& \left. - i\left(-\rho\int_0^t y_0 e^{\nu m_s} dB_s + \frac{y_0^2}{2}\int_0^t e^{2\nu m_s} ds\right)\xi + \frac{\bar{\rho}^2 y_0^2 \xi^2}{2}\int_0^t e^{2\nu m_s} ds \right| \\
\le\ & 2|\mathcal{R}_t + i\mathcal{I}_t|
\end{aligned} \tag{A4}$$

since $e^{-\frac{\bar{\rho}^2 y_0^2 v_t}{2}\xi^2} + e^{-\frac{\bar{\rho}^2 y_0^2 \xi^2}{2}\int_0^t e^{2\nu m_s} ds} \le 2$ for all t and ξ. Apparently, $\mathcal{R}_t$ and $\mathcal{I}_t$ are given by

$$\begin{aligned}
\mathcal{R}_t &= \left[-v_t + \int_0^t e^{2\nu m_s} ds\right]\frac{\bar{\rho}^2 y_0^2 \xi^2}{2}, \\
\mathcal{I}_t &= \left(-\rho\int_0^t y_0 e^{\nu B_s^H} dB_s + \frac{y_0^2 v_t}{2} + \rho\int_0^t y_0 e^{\nu m_s} dB_s - \frac{y_0^2}{2}\int_0^t e^{2\nu m_s} ds\right)\xi.
\end{aligned}$$

In the following, K denotes a generic constant whose value may vary in different contexts. Then, by (A3) and (A4) and Hölder's inequality, we have

$$\begin{aligned}
&\left|\frac{1}{2\pi}\int e^{-ix_0\xi}\mathbb{E}\left[\hat{f}(\xi, Y_t)(\lambda_1(t) - \lambda_2(t))\right]d\xi\right| \\
\le\ & 2\int \mathbb{E}\left[|\hat{f}(\xi, Y_t)||\mathcal{R}_t + i\mathcal{I}_t|\right]d\xi \\
\le\ & 2\left(\mathbb{E}\int |\hat{f}(\xi, Y_t)|^{(1-\epsilon)p} d\xi\right)^{\frac{1}{p}}\left(\int \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q}|\mathcal{R}_t + i\mathcal{I}_t|^q\right]d\xi\right)^{\frac{1}{q}} \\
\le\ & K\left(\mathbb{E}\int |\hat{f}(\xi, Y_t)|^{(1-\epsilon)p} d\xi\right)^{\frac{1}{p}}\left(\int \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q}(|\mathcal{R}_t|^q + |\mathcal{I}_t|^q)\right]d\xi\right)^{\frac{1}{q}},
\end{aligned} \tag{A5}$$

for some $\epsilon \in (0, 1)$ and $\frac{1}{p} + \frac{1}{q} = 1$, $p, q > 0$.

Since $f \in C_0^\infty(\mathbb{R}^2)$, it is easy to show the following properties of $\hat{f}$:

(i) for any $r \geq 0$, $\sup_{(\xi,y)\in\mathbb{R}^2} \left|\xi^r \hat{f}(\xi,y)\right| < \infty$;

(ii) for any $r \geq 0$ and $p > 0$, $\int |\xi|^r \sup_{y\in\mathbb{R}} |\hat{f}(\xi,y)|^p d\xi < \infty$.

Note that property (ii) can be easily obtained by property (i).

By the above property (ii), we can show that

$$\limsup_{t\to 0^+} \mathbb{E} \int |\hat{f}(\xi,Y_t)|^{(1-\epsilon)p} d\xi < \infty. \tag{A6}$$

We compute the second term in (A5) separately as follows. By changing variables, we get

$$\begin{aligned}
& \int \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q} |\mathcal{R}_t|^q\right] d\xi \\
\leq\ & K\bar{\rho}^{2q} y_0^{2q} \int \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q}\left(v_t^q + \left(\int_0^t e^{2\nu m_s} ds\right)^q\right)\right]\xi^{2q} d\xi \\
=\ & K\bar{\rho}^{2q} y_0^{2q} t^q \int \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q}\left(\left(\int_0^1 e^{2\nu B_{tu}^H} du\right)^q + \left(\int_0^t e^{2R(1,u)\eta} du\right)^q\right)\right]\xi^{2q} d\xi \\
=\ & K\bar{\rho}^{2q} y_0^{2q} t^q (L_1 + L_2),
\end{aligned} \tag{A7}$$

where

$$\begin{aligned}
L_1 &:= \int \xi^{2q} \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q}\left(\int_0^1 e^{2\nu B_{tu}^H} du\right)^q\right] d\xi, \\
L_2 &:= \int \xi^{2q} \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q}\left(\int_0^1 e^{2R(1,u)\eta} du\right)^q\right] d\xi.
\end{aligned}$$

By property (ii) of $\hat{f}$, it is easy to see that

$$\limsup_{t\to 0^+} L_2 \leq \left(\int_0^1 e^{2R(1,u)\eta} du\right)^q \int \xi^{2q} \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q}\right] d\xi < \infty. \tag{A8}$$

For L_1, by Jensen's inequality and Hölder's inequality, we have

$$\begin{aligned}
L_1 &\leq \int \xi^{2q} \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q} \int_0^1 e^{2q\nu B_{tu}^H} du\right] d\xi \\
&\leq \int \xi^{2q} \left\{\mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q p_1}\right]\right\}^{\frac{1}{p_1}} \left\{\mathbb{E}\left[\left(\int_0^1 e^{2q\nu B_{tu}^H} du\right)^{q_1}\right]\right\}^{\frac{1}{q_1}} d\xi \\
&\leq \int \xi^{2q} \left\{\mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q p_1}\right]\right\}^{\frac{1}{p_1}} \left\{\int_0^1 \mathbb{E}\left[e^{2q q_1 \nu B_{tu}^H}\right] du\right\}^{\frac{1}{q_1}} d\xi \\
&= \int \xi^{2q} \left\{\mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q p_1}\right]\right\}^{\frac{1}{p_1}} \left\{\int_0^1 e^{2(q q_1 \nu)^2 (tu)^{2H}} du\right\}^{\frac{1}{q_1}} d\xi.
\end{aligned}$$

where $\frac{1}{p_1} + \frac{1}{q_1} = 1$ with $p_1, q_1 > 0$. Therefore, using property (ii) again, we can easily show

$$\limsup_{t\to 0^+} L_1 \leq \limsup_{t\to 0^+} \int \xi^{2q} \left\{\mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q p_1}\right]\right\}^{\frac{1}{p_1}} d\xi \left\{\int_0^1 e^{2(q q_1 \nu)^2 (u)^{2H}} du\right\}^{\frac{1}{q_1}} < \infty. \tag{A9}$$

Thus, it implies from (A7)–(A9) that

$$\int \mathbb{E}\left[|\hat{f}(\xi,Y_t)|^{\epsilon q} |\mathcal{R}_t|^q\right] d\xi = O(t^q), \tag{A10}$$

for any $q > 1$, as $t \to 0^+$.

Similarly, we can write

$$
\begin{aligned}
& \int \mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q} |\mathcal{I}_t|^q\Big] d\xi \\
\leq \; & K \int |\xi|^q \mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q} \times \\
& \left\{ \left|\rho \int_0^t y_0 e^{\nu B_s^H} dB_s\right|^q + \left|\frac{y_0^2 v_t}{2}\right|^q + \left|\rho \int_0^t y_0 e^{\nu m_s} dB_s\right|^q + \left|\frac{y_0^2}{2} \int_0^t e^{2\nu m_s} ds\right|^q \right\}\Bigg] d\xi \\
= \; & K(J_1 + J_2 + J_3 + J_4),
\end{aligned}
$$

where

$$
\begin{aligned}
J_1 &:= |\rho|^q \int |\xi|^q \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q} \left|\int_0^t y_0 e^{\nu B_s^H} dB_s\right|^q\right] d\xi, \\
J_2 &:= \int |\xi|^q \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q} \left|\frac{y_0^2 v_t}{2}\right|^q\right] d\xi, \\
J_3 &:= \int |\xi|^q \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q} \left|\rho \int_0^t y_0 e^{\nu m_s} dB_s\right|^q\right] d\xi, \\
J_4 &:= \int |\xi|^q \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q} \left|\frac{y_0^2}{2} \int_0^t e^{2\nu m_s} ds\right|^q\right] d\xi.
\end{aligned}
$$

We estimate J_1 through J_4 separately as follows.

- J_1: Choosing $p_1 > 0$ such that $\frac{qq_1}{2} > 1$, by Hölder's inequality, the Burkholder-Davis-Gundy inequality, Jensen's inequality and a change of variables, we obtain Notice that

$$
\begin{aligned}
J_1 &\leq |\rho|^q y_0^q \int |\xi|^q \left\{\mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1}\Big]\right\}^{\frac{1}{p_1}} \left\{\mathbb{E}\left[\left|\int_0^t e^{\nu B_s^H} dB_s\right|^{qq_1}\right]\right\}^{\frac{1}{q_1}} d\xi \\
&\leq |\rho|^q y_0^q \int |\xi|^q \left\{\mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1}\Big]\right\}^{\frac{1}{p_1}} \left\{\mathbb{E}\left[\left|\int_0^t e^{2\nu B_s^H} ds\right|^{\frac{qq_1}{2}}\right]\right\}^{\frac{1}{q_1}} d\xi \\
&= |\rho|^q y_0^q t^{\frac{q}{2}} \int |\xi|^q \left\{\mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1}\Big]\right\}^{\frac{1}{p_1}} \left\{\mathbb{E}\left[\left|\int_0^1 e^{2\nu B_{tu}^H} du\right|^{\frac{qq_1}{2}}\right]\right\}^{\frac{1}{q_1}} d\xi \\
&\leq |\rho|^q y_0^q t^{\frac{q}{2}} \int |\xi|^q \left\{\mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1}\Big]\right\}^{\frac{1}{p_1}} \left\{\int_0^1 \mathbb{E}\Big[e^{qq_1 \nu B_{tu}^H}\Big] du\right\}^{\frac{1}{q_1}} d\xi \\
&= |\rho|^q y_0^q t^{\frac{q}{2}} \int |\xi|^q \left\{\mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1}\Big]\right\}^{\frac{1}{p_1}} \left\{\int_0^1 e^{\frac{(qq_1\nu)^2}{2}(tu)^{2H}} du\right\}^{\frac{1}{q_1}} d\xi.
\end{aligned}
$$

  By property (ii) we have

$$
\begin{aligned}
& \limsup_{t \to 0^+} \int |\xi|^q \left\{\mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1}\Big]\right\}^{\frac{1}{p_1}} \left\{\int_0^1 e^{\frac{(qq_1\nu)^2}{2}(tu)^{2H}} du\right\}^{\frac{1}{q_1}} d\xi \\
\leq \; & \limsup_{t \to 0^+} \int \xi^{2q} \left\{\mathbb{E}\Big[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1}\Big]\right\}^{\frac{1}{p_1}} d\xi \left\{\int_0^1 e^{2(qq_1\nu)^2 (u)^{2H}} du\right\}^{\frac{1}{q_1}} < \infty.
\end{aligned}
$$

  Thus, we can see that $J_1 = O(t^{\frac{q}{2}})$ as $t \to 0^+$.
- J_2 and J_4: The asymptotic behavior of J_2 and J_4 is the same as that of $t^q L_1$, and hence, $J_2, J_4 = O(t^q)$ as $t \to 0^+$.
- J_3: By using the same technique to J_1, we have

$$J_3 \leq |\rho|^q y_0^q t^{\frac{q}{2}} \int |\xi|^q \left\{ \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q p_1} \right] \right\}^{\frac{1}{p_1}} \left\{ \mathbb{E}\left[\left| \int_0^1 e^{2R(1,u)\eta} du \right|^{\frac{q q_1}{2}} \right] \right\}^{\frac{1}{q_1}} d\xi$$

and $J_3 = O(t^{\frac{q}{2}})$ as $t \to 0^+$.

Thus, putting all the estimates for the J_i's together we get

$$\int \mathbb{E}\left[|\hat{f}(\xi, Y_t)|^{\epsilon q} |\mathcal{I}_t|^q \right] d\xi = O(t^{\frac{q}{2}}), \tag{A11}$$

for any $q > 1$, as $t \to 0^+$.

Therefore, it implies from (A5), (A6), (A10) and (A11) that

$$\left| \frac{1}{2\pi} \int e^{-ix_0\xi} \mathbb{E}\left[\hat{f}(\xi, Y_t)(\lambda_1(t) - \lambda_2(t)) \right] d\xi \right| = O(t^{\frac{1}{2}}),$$

that is, (A3) converges to zero in the order of $t^{\frac{1}{2}}$ as $t \to 0$, for every $f \in C_0^\infty(\mathbb{R}^2)$.

Appendix A.2. Laplace Asymptotic Formula

We prove the following form of the Laplace asymptotic formula required in the derivation of the small time asymptotic of the price of an out-of-money call.

Lemma A1 (Laplace asymptotic formula). *Let $\mathcal{C}$ be a closed and convex set in $\mathbb{R}^2$ with a nonempty and smooth boundary $\partial\mathcal{C}$. Suppose that $\theta(t,x) := \theta_0(x) + t^\alpha \theta_1(x) + t^{2\alpha}\theta_2(x)$, with $0 \leq 2\alpha < 1$, has continuous second-order partial derivatives in $x \in \mathcal{C}$, and, for every t sufficiently small, the function $\theta(t,x)$ is locally convex in $\mathcal{C}$ and attains its minimum uniquely at $x^*(t) \in \partial\mathcal{C}$. Moreover, there is $\epsilon_0 > 0$ such that for any $0 < \epsilon < \epsilon_0$, there exist t_0 and $\delta > 0$ for which*

$$\theta(t,x) \geq \theta(t, x^*(t)) + \delta, \ \forall (t,x) \in [0,t_0] \times (\mathcal{C} \setminus B_\epsilon(x^*(t))),$$

where $B_\epsilon(x^*(t)) = \{x : |x - x^*(t)| < \epsilon\}$ is the open ball of radius ϵ centered at $x^*(t)$.

Assume that f has continuous second-order partial derivatives in $\mathcal{C}$, is integrable over $\mathcal{C}$ (i.e., $\int_{\mathcal{C}} |f(x)|dx < \infty$) and that f vanishes identically in $\mathcal{C}^c$ and on the boundary $\partial\mathcal{C}$ but the inward normal directional derivative of f at $x^*(t)$ is nonzero.

Then, we have the asymptotic expansion, as $t \to 0^+$,

$$\begin{aligned} &\int_{\mathcal{C}} e^{-\frac{\theta(x,t)}{t}} f(x)dx \\ = \ &\frac{\sqrt{2\pi}\, t^{\frac{5}{2}}\, e^{-\frac{\theta(t,x^*(t))}{t}}}{\sqrt{\partial^2_{\mathbf{tan}}\theta(t,x^*(t))}|\nabla\theta(t,x^*(t))|} \left[\frac{\nabla f(x^*(t)) \cdot \nabla\theta(t,x^*(t))}{|\nabla\theta(t,x^*(t))|^2} + \frac{1}{2} \frac{\partial^2_{\mathbf{tan}} f(x^*(t))}{\partial^2_{\mathbf{tan}}\theta(t,x^*(t))} + o(1) \right], \end{aligned} \tag{A12}$$

where $\partial^2_{\mathbf{tan}} f(x^*)$ and $\partial^2_{\mathbf{tan}}\theta(t,x^*)$ are the second derivatives of f and θ respectively in the tangential direction to $\mathcal{C}$ at x^*.

Proof. For any $0 < \epsilon < \epsilon_0$, we split the integral on the left side of (A12) into two parts as

$$\int_{\mathcal{C}} e^{-\frac{\theta(t,x)}{t}} f(x)dx = \int_{\mathcal{C} \bigcap B_\epsilon(x^*(t))} e^{-\frac{\theta(t,x)}{t}} f(x)dx + \int_{\mathcal{C} \setminus B_\epsilon(x^*(t))} e^{-\frac{\theta(t,x)}{t}} f(x)dx. \tag{A13}$$

We treat the two terms on the right-hand side of (A13) individually. For the first term, since the integration region is restricted to a subset of the small ball $B_\epsilon(x^*(t))$, it can be reparametrized by $y = (y^1, y^2)$ so that in the y-coordinates the set $\{y : y^2 = 0\}$ corresponds to $\partial\mathcal{C}$ and the vectors $\{\partial_{y^1}, \partial_{y^2}\}$ form a local orthonormal frame around $x^*(t)$. For simplicity, we further assume that in the y-coordinates $x^*(t)$ is located at the origin. Note that in the y-coordinates the vector ∂_{y^2} is parallel to $\nabla\theta(x^*(t))$ as well as the inward normal vector of $\mathcal{C}$ at $x^*(t)$.

We shall use the convention that repeated indices are summed up over their respective ranges. Denote partial derivatives by subindices, we have for $y \in B_\epsilon(x^*(t))$

$$\theta(t,y) = \theta(t,0) + \theta_2(t,0)y^2 + \frac{1}{2}\theta_{ij}(t,0)y^i y^j + o(|y|^2),$$
$$f(y) = f_i(0)y^i + \frac{1}{2}f_{ij}(0)y^i y^j + o(|y|^2)$$

since $\theta_1(0) = 0$ for θ attains its minimum at the boundary point $x^*(t)$.

Thus, in the y-coordinates the first integral on the right-hand side of (A13) reads

$$\begin{aligned} &\int_{\mathcal{C}\bigcap B_\epsilon(x^*(t))} e^{-\frac{\theta(t,x)}{t}} f(x)dx \\ \approx\ &\int_0^\epsilon \int_{-\epsilon}^\epsilon e^{-\frac{1}{t}\left(\theta(t,0)+\theta_2(t,0)y^2+\frac{1}{2}\theta_{ij}(t,0)y^i y^j\right)} \left[f_i(0)y^i + \frac{1}{2}f_{ij}(0)y^i y^j\right] dy^1 dy^2. \end{aligned} \tag{A14}$$

Now, by a change of variables

$$y^1 = \sqrt{t}z^1, \quad y^2 = tz^2,$$

we can write the above integral on the right-hand side of (A14) as

$$\begin{aligned} &e^{-\frac{\theta(t,0)}{t}} t^{\frac{3}{2}} \int_0^{\frac{\epsilon}{t}} \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\theta_2(t,0)z^2+\frac{1}{2}\theta_{11}(t,0)(z^1)^2+\theta_{12}(t,0)z^1z^2\sqrt{t}+\frac{1}{2}\theta_{22}(t,0)(z^2)^2 t\right)} \times \\ &\left[f_1(0)z^1\sqrt{t} + f_2(0)z^2 t + \frac{1}{2}f_{11}(0)(z^1)^2 t + f_{12}(0)z^1z^2 t^{\frac{3}{2}} + \frac{1}{2}f_{22}(0)(z^2)^2 t^2\right] dz^1 dz^2. \end{aligned} \tag{A15}$$

Note that, for any real numbers $a_1, \ldots, a_5$, by dominated convergence theorem, we have

$$\begin{aligned} &\lim_{t\to 0}\int_0^{\frac{\epsilon}{t}} \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\theta_2(t,0)z^2+\frac{1}{2}\theta_{11}(t,0)(z^1)^2+\theta_{12}(t,0)z^1z^2\sqrt{t}+\frac{1}{2}\theta_{22}(t,0)(z^2)^2 t\right)} \times \\ &\left[a_1z^1 + a_2z^2 + a_3(z^1)^2 + a_4(z^2)^2 + a_5z^1z^2\right] dz^1dz^2 \\ = &\int_0^\infty \int_{-\infty}^\infty e^{-\left(\theta_2(0,0)z^2+\frac{1}{2}\theta_{11}(0,0)(z^1)^2\right)} \times \\ &\left[a_1z^1 + a_2z^2 + a_3(z^1)^2 + a_4(z^2)^2 + a_5z^1z^2\right] dz^1dz^2 \in (-\infty, \infty). \end{aligned}$$

Thus, the quantity in (A15) equals

$$\begin{aligned} &e^{-\frac{\theta(t,0)}{t}} t^{\frac{3}{2}} \left\{ \int_0^{\frac{\epsilon}{t}} \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\theta_2(t,0)z^2+\frac{1}{2}\theta_{11}(t,0)(z^1)^2\right)} \times \right. \\ &\left. \left[f_1(0)z^1\sqrt{t} + f_2(0)z^2 t + \frac{1}{2}f_{11}(0)(z^1)^2 t\right] dz^1dz^2 + O\left(t^{\frac{1}{2}}\right) \right\} \\ = \ &e^{-\frac{\theta(t,0)}{t}} t^{\frac{3}{2}} \left[\sqrt{t}\cdot I + t\cdot II + t\cdot III + O\left(t^{\frac{1}{2}}\right)\right], \end{aligned} \tag{A16}$$

where

$$I = \int_0^{\frac{\epsilon}{t}} \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\theta_2(t,0)z^2+\frac{1}{2}\theta_{11}(t,0)(z^1)^2\right)} f_1(0)z^1 dz^1dz^2,$$
$$II = \int_0^{\frac{\epsilon}{t}} \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\theta_2(t,0)z^2+\frac{1}{2}\theta_{11}(t,0)(z^1)^2\right)} f_2(0)z^2 dz^1dz^2,$$
$$III = \frac{1}{2}\int_0^{\frac{\epsilon}{t}} \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\theta_2(t,0)z^2+\frac{1}{2}\theta_{11}(t,0)(z^1)^2\right)} f_{11}(0)(z^1)^2 dz^1dz^2.$$

As $t \to 0^+$, we calculate each integral individually as follows. For I, since the function in z^1 is an odd function and the integral interval for z^1 is symmetric about the origin, we obtain

$$\begin{aligned} I &= f_1(0) \int_0^{\frac{\epsilon}{t}} e^{-\theta_2(t,0)z^2} dz^2 \times \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\frac{1}{2}\theta_{11}(t,0)(z^1)^2\right)} z^1 dz^1 \\ &= 0. \end{aligned} \tag{A17}$$

For II and III, notice that $\theta_2(t,0) > 0$ and $\theta_{11}(t,0) > 0$, and hence, we obtain

$$\begin{aligned} II &= f_2(0) \int_0^{\frac{\epsilon}{t}} e^{-\theta_2(t,0)z^2} z^2 dz^2 \times \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\frac{1}{2}\theta_{11}(t,0)(z^1)^2\right)} dz^1 \\ &\approx f_2(0) \int_0^{\infty} e^{-\theta_2(t,0)z^2} z^2 dz^2 \times \int_{-\infty}^{\infty} e^{-\frac{1}{2}\theta_{11}(t,0)(z^1)^2} dz^1 \\ &= \frac{f_2(0)}{\theta_2^2(t,0)} \times \sqrt{\frac{2\pi}{\theta_{11}(t,0)}}, \end{aligned} \tag{A18}$$

and

$$\begin{aligned} III &= \frac{f_{11}(0)}{2} \int_0^{\frac{\epsilon}{t}} e^{-\theta_2(t,0)z^2} dz^2 \times \int_{-\frac{\epsilon}{\sqrt{t}}}^{\frac{\epsilon}{\sqrt{t}}} e^{-\left(\frac{1}{2}\theta_{11}(t,0)(z^1)^2\right)} (z^1)^2 dz^1 \\ &\approx \frac{f_{11}(0)}{2} \int_0^{\infty} e^{-\theta_2(t,0)z^2} dz^2 \times \int_{-\infty}^{\infty} e^{-\frac{1}{2}\theta_{11}(t,0)(z^1)^2} (z^1)^2 dz^1 \\ &= \frac{f_{11}(0)}{2\theta_2(t,0)} \times \sqrt{\frac{2\pi}{\theta_{11}^3(t,0)}}. \end{aligned} \tag{A19}$$

Therefore, it implies from (A14)–(A19) that, in the y-coordinates,

$$\int_{\mathcal{C} \cap B_\epsilon(x^*(t))} e^{-\frac{\theta(t,x)}{t}} f(x) dx \approx e^{-\frac{\theta(t,0)}{t}} t^{\frac{5}{2}} \sqrt{\frac{2\pi}{\theta_{11}(t,0)}} \left[\frac{f_2(0)}{\theta_2^2(t,0)} + \frac{f_{11}(0)}{2\theta_2(t,0)\theta_{11}(t,0)} + o(1) \right]. \tag{A20}$$

For the second term on the right-hand side of (A13), we get

$$\left| \int_{\mathcal{C} \setminus B_\epsilon(x^*)} e^{-\frac{\theta(t,x)}{t}} f(x) dx \right| \le \int_{\mathcal{C} \setminus B_\epsilon(x^*)} e^{-\frac{\theta(t,x^*)+\delta}{t}} |f(x)| dx \le e^{-\frac{\delta}{t}} e^{-\frac{\theta(t,x^*)}{t}} \int_{\mathcal{C}} |f(x)| dx. \tag{A21}$$

As a result, the second term is exponentially small (at the rate δ) as $t \to 0^+$ compared to the expansion (A12) obtained for the first term, hence it does not contribute to the asymptotic expansion.

Finally, by (A13), (A20) and (A21) we obtain the Laplace expansion (A12) by rewriting the expressions for the right-hand side of (A20) in the x-coordinates. □

References

Baillie, Richard T., Tim Bollerslev, and Hans Ole Mikkelsen. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74: 3–30. [CrossRef]

Cheng, Xue, and Tai-Ho Wang. 2018. Bessel bridge representation for heat kernel in hyperbolic space. *Proceedings of the American Mathematical Society* 146: 1781–92. [CrossRef]

Comte, Fabienne, and Eric Renault. 1998. Long memory in continuous-time stochastic volatility models. *Mathematical Finance* 8: 291–323. [CrossRef]

Comte, Fabienne, Laure Coutin, and Eric Renault. 2012. Affine fractional stochastic volatility models. *Annals of Finance* 8: 337–78. [CrossRef]

Cont, Rama, and Purba Das. 2022. Rough Volatility: Fact or Artifact? *arXiv* arXiv:2203.13820.

Decreusefond, Laurent, and A. Suleyman Üstünel. 1999. Stochastic analysis of the fractional Brownian motion. *Potential Analysis* 10: 177–214. [CrossRef]

Ekström, Erik, and Bing Lu. 2015. Short-time implied volatility in exponential Lévy models. *International Journal of Theoretical and Applied Finance* 18: 1550025. [CrossRef]

El Euch, Omar, and Mathieu Rosenbaum. 2019. The characteristic function of rough Heston models. *Mathematical Finance* 29: 3–38. [CrossRef]

Gao, Kun, and Roger Lee. 2014. Asymptotics of implied volatility to arbitrary order. *Finance and Stochastics* 18: 349–92. [CrossRef]

Gatheral, Jim, Thibault Jaisson, and Mathieu Rosenbaum. 2018. Volatility is rough. *Quantitative Finance* 18: 933–49. [CrossRef]

Granger, Clive W. J., and Roselyne Joyeux. 1980. An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis* 1: 15–39. [CrossRef]

Guennoun, Hamza, Antoine Jaquier, Patrick Roome, and Fangwei Shi. 2018. Asymptotic behavior of the fractional Heston model. *SIAM Journal on Financial Mathematics* 9: 337–78. [CrossRef]

Hagan, Patrick, Deep Kumar, Andrew Lesniewski, and Diana Woodward. 2002. Managing smile risk. *Wilmott Magazine* 1: 84–108.

Hagan, Patrick, Andrew Lesniewski, and Diana Woodward. 2015. Probability distribution in the SABR Model of stochastic volatility. In *Large Deviations and Asymptotic Methods in Finance*. Springer Proceedings in Mathematics & Statistics. Cham: Springer, vol. 110, pp. 1–35.

Hu, Yaozhong. 2017. *Analysis on Gaussian Space*. Singapore: World Scientific.

Ikeda, Nobuyuki, and Hiroyuki Matsumoto. 1999. Brownian Motion on the Hyperbolic Plane and Selberg Trace Formula. *Journal of Functional Analysis* 163: 63–100. [CrossRef]

Jarrow, Robert A., Philip Protter, and Hasanjan Sayit. 2009. No arbitrage without semimartingales. *The Annals of Applied Probability* 19: 596–616. [CrossRef]

Mishura, Yuliya. 2008. *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. Lecture Notes in Mathematics. Berlin and Heidelberg: Springer Science & Business Media, vol. 1929.

Nualart, David. 2006. *The Malliavin Calculus and Related Topics*, 2nd ed. Berlin and Heidelberg: Springer.

Rogers, Chris. 2019. Things We Think We Know. Available online: https://www.skokholm.co.uk/wp-content/uploads/2019/11/TWTWKpaper.pdf (accessed on 1 June 2022).

Roper, Michael, and Marek Rutkowski. 2009. A note on the behaviour of the Black-Scholes implied volatility close to expiry. *International Journal of Theoretical and Applied Finance* 12: 427–41. [CrossRef]