Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Short-time near-the-money skew in rough fractional volatility models

C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath & B. Stemper

To cite this article: C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath & B. Stemper (2019) Short-time near-the-money skew in rough fractional volatility models, Quantitative Finance, 19:5, 779-798, DOI: 10.1080/14697688.2018.1529420

To link to this article: https://doi.org/10.1080/14697688.2018.1529420

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 13 Nov 2018.

Submit your article to this journal

Article views: 1995

View related articles

View Crossmark data

Citing articles: 24 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rquf20

Quantitative Finance, 2019
Vol. 19, No. 5, 779–798, https://doi.org/10.1080/14697688.2018.1529420

Routledge
Taylor & Francis Group
Check for updates

Short-time near-the-money skew in rough fractional volatility models

C. BAYER†, P. K. FRIZ†‡*, A. GULISASHVILI§, B. HORVATH¶ and B. STEMPER†‡

†WIAS Berlin, Berlin, Germany
‡TU Berlin, Berlin, Germany
§Ohio University, Athens, USA
¶Imperial College London, London, UK

(*Received 22 March 2017; accepted 7 September 2018; published online 13 November 2018*)

We consider rough stochastic volatility models where the driving noise of volatility has fractional scaling, in the 'rough' regime of Hurst parameter $H < 1/2$. This regime recently attracted a lot of attention both from the statistical and option pricing point of view. With focus on the latter, we sharpen the large deviation results of Forde-Zhang [Asymptotics for rough stochastic volatility models. *SIAM J. Financ. Math.*, 2017, **8**(1), 114–145] in a way that allows us to zoom-in around the money while maintaining full analytical tractability. More precisely, this amounts to proving higher order moderate deviation estimates, only recently introduced in the option pricing context. This in turn allows us to push the applicability range of known at-the-money skew approximation formulae from CLT type log-moneyness deviations of order $t^{1/2}$ (works of Alòs, León & Vives and Fukasawa) to the wider moderate deviations regime.

Keywords: Rough stochastic volatility model; European option pricing; Small-time asymptotics; Moderate deviations

2010 Mathematics Subject Classification: 91G20, 60H30, 60F10, 60H07, 60G22, 60G18

1. Introduction

Since the groundbreaking work of Gatheral *et al*. (2014), the past two years have brought about a gradual shift in volatility modeling, leading away from classical diffusive stochastic volatility models towards so-called rough volatility models. The term was coined in Gatheral *et al*. (2014) and Bayer *et al*. (2016), and it essentially describes a family of (continuous-path) stochastic volatility models where the driving noise of the volatility process has Hölder regularity lower than Brownian motion, typically achieved by modeling the fundamental noise innovations of the volatility process as a fractional Brownian motion with Hurst exponent (and hence Hölder regularity) $H < 1/2$. Here, we would also like to mention pioneering work on asymptotics for rough volatility models in Alòs *et al*. (2007) and Fukasawa (2011). A major appeal of such rough volatility models lies in the fact that they effectively capture several stylized facts of financial markets both from a statistical (Gatheral *et al*. 2014; Bennedsen *et al*. 2016) and an option-pricing point of view (Bayer *et al*. 2016). In particular, with regards to the latter point of view, a widely observed empirical phenomenon in equity markets is the 'steepness of the smile on the short end' describing the fact that as time to maturity becomes small the empirical implied volatility skew follows a power law with negative exponent, and thus becomes arbitrarily large near zero. While standard stochastic volatility models with continuous paths struggle to capture this phenomenon, predicting instead a constant at-the-money implied volatility behavior on the short end (Gatheral 2011), models in the fractional stochastic volatility family (and more specifically so-called rough volatility models) constitute a class, well-tailored to fit empirical implied volatilities for short dated options.

Typically, the popularity of asset pricing models hinges on the availability of efficient numerical pricing methods. In the case of diffusions, these include Monte Carlo estimators, PDE discretization schemes, asymptotic expansions and transform methods. With fractional Brownian motion being the prime example of a process beyond the semimartingale framework, most currently prevalent option pricing methods

*Corresponding author. Email: friz@math.tu-berlin.de, P.K.Friz@gmail.com

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

– particularly the ones assuming semimartingality or Markovianity – may not easily carry over to the rough setting. In fact, the memory property (aka non-Markovianity) of fractional Brownian motion rules out PDE methods, heat kernel methods and all related methods involving a Feynman-Kac-type Ansatz. Previous work has thus focused on finding efficient Monte Carlo simulation schemes (Bayer *et al*. 2016; Bennedsen *et al*. 2017; Bayer *et al*. 2017) or – in the special case of the Rough Heston model – on an explicit formula for the characteristic function of the log-price (see El Euch and Rosenbaum 2016), thus in this particular model making pricing amenable to Fourier based methods. In our work, we rely on small-maturity approximations of option prices. This is a well-studied topic for which we mention (with no claim to completeness) a number of works, either based on large deviations or central limit type scaling regime, that inspired this work: Alòs *et al*. (2007), Fukasawa (2011), Deuschel *et al*. (2014a), Deuschel *et al*. (2014b) and Fukasawa (2017), also Medvedev and Scaillet (2003, 2007), Osajima (2007, 2015), Guennoun *et al*. (2014), Mijatović and Tankov (2016) and especially Forde and Zhang (2017). Rather recently, Friz *et al*. (2018) introduced another regime called moderately-out-of-the-money (MOTM), which, in a sense, effectively navigates between the two regimes mentioned above, by rescaling the strike with respect to the time to maturity. This approach has various advantages. On the one hand, it reflects the market reality that as time to maturity approaches zero, strikes with acceptable bid-ask spreads tend to move closer to the money (see Friz *et al*. 2018 for more details). On the other hand, it allows us to zoom in on the term structure of implied volatility around the money at a high resolution scale. To be more specific, our paper adds to the existing literature in two ways. First, we obtain a generalization of the Osajima energy expansion (Osajima 2015) to a non-Markovian case, and using the new expansion, we extend the analysis of Friz *et al*. (2018) to the case, where the volatility is driven by a rough ($H < 1/2$) fractional Brownian motion. Indeed, Laplace approximation methods on Wiener space in the spirit of Azencott (1982, 1985), Ben Arous (1988) and Bismut (1984) can be adapted to the present context, so that our analysis builds upon this framework in a fractional setting. Unlike many other works in this field, we do not rely on density expansions. Finally, using a version of the '*rough Bergomi model*' (Bayer *et al*. 2016), we demonstrate numerically that our implied volatility asymptotics capture very well the geometry of the term structure of implied volatility over a wide array of maturities, extending up to a year.

The paper is organized as follows: In Section 2 we set the scene, describing the class of models included in our framework ((1) and (2)) and recalling some known results ((4) and (7)), which are the starting point of our analysis. Most importantly, we argue that for small-time considerations it would suffice to restrict our attention to a class of stochastic volatility models of the form (3) with a volatility process driven by a Gaussian Volterra process such as in (2). We formulate general assumptions on the Volterra kernel (Assumptions 2.1 and 2.5) and on the function σ in (3) (Assumption 2.4) under which our results are valid. In Section 3 we gather our main results, concerning a higher order expansion of the energy (Theorem 3.1), and a general expansion formula for the corresponding call prices. We derive the classical Black-Scholes expansion for the call price, using the latter result mentioned above. In addition, in Section 3 we formulate moderate deviation expansions, which allow us to derive the corresponding asymptotic formulae for implied volatilities and implied volatility skews. Finally, Section 4 displays our simulation results. Sections 5–7 are devoted to proofs of the energy expansion, the price expansion and the moderate deviations expansion, respectively. In the appendix, we have collected some auxiliary lemmas, which are used in different sections.

2. Exposition and assumptions

We consider a rough stochastic volatility model, normalized to $r = 0$ and $S_0 = 1$, of the form suggested by Forde and Zhang (2017)

$$\frac{\mathrm{d}S_t}{S_t} = \sigma(\hat{B}_t)\,\mathrm{d}\left(\bar{\rho}W_t + \rho B_t\right). \tag{1}$$

Here (W, B) are two independent standard Brownian motions, $\rho \in (-1, 1)$ a correlation parameter, and $\bar{\rho}^2 = 1 - \rho^2$. Then $\bar{\rho}W + \rho B$ is another standard Brownian motion which has constant correlation ρ with the factor B, which drives the stochastic volatility

$$\sigma_{\text{stoch}}(t, \omega) := \sigma(\hat{B}_t(\omega)) \equiv \sigma(\hat{B}).$$

Here $\sigma(.)$ is some real-valued function, typically smooth but not bounded, and we will denote by $\sigma_0 := \sigma(0)$ the spot volatility, with $\hat{B}$ a Gaussian (Volterra) process of the form

$$\hat{B}_t = \int_0^t K(t, s)\,\mathrm{d}B_s, \quad t \geq 0, \tag{2}$$

for some kernel K, which shall be further specified in Assumptions 2.1 and 2.5 below. The log-price $X_t = \log(S_t)$ satisfies

$$\mathrm{d}X_t = -\tfrac{1}{2}\sigma^2(\hat{B}_t)\,\mathrm{d}t + \sigma(\hat{B}_t)\,\mathrm{d}\left(\bar{\rho}W + \rho B\right), \quad X_0 = 0. \tag{3}$$

Recall that by Brownian scaling, for fixed $t > 0$,

$$(B_{ts}, W_{ts})_{s\geq 0} \overset{law}{=} \varepsilon(B_s, W_s)_{s\geq 0}, \quad \text{where } \varepsilon \equiv \varepsilon(t) \equiv t^{1/2}.$$

As a direct consequence, classical short-time SDE problems can be analyzed as small-noise problems on a unit time horizon. For our analysis, it will also be crucial to impose such a scaling property on the Gaussian process $\hat{B}$ (more precisely, on the kernel K in (2)) driving the volatility process in our model:

Assumption 2.1 Small time self-similarity *There exists a number t_0 with $0 < t_0 \leq 1$ and a function $t \mapsto \hat{\varepsilon} = \hat{\varepsilon}(t)$, $0 \leq t \leq t_0$, such that*

$$(\hat{B}_{ts} : 0 \leq s \leq t_0) \overset{law}{=} (\hat{\varepsilon}\hat{B}_s : 0 \leq s \leq t_0).$$

In fact, we will always have

$$\hat{\varepsilon} \equiv \hat{\varepsilon}(t) \equiv t^H = \varepsilon^{2H},$$

which covers the examples of interest, in particular standard fractional Brownian motion $\hat{B} = B^H$ or Riemann-Liouville fBM with explicit kernel $K(t,s) = \sqrt{2H}\,|t-s|^{H-1/2}$. (This is very natural, even from a general perspective of self-similar processes, see Lamperti 1962.)

We insist that no (global) self-similarity of $\hat{B}$ is required, as only $\hat{B}|_{[0,t]}$ for arbitrarily small t matters.

REMARK 2.2 It should be possible to replace the fractional Brownian motion by a certain fractional Ornstein-Uhlenbeck process in the results obtained in this paper. Intuitively, this replacement creates a negligible perturbation (for $t \ll 1$) of the fBm environment. A similar situation was in fact encountered in Cass and Friz (2010), where fractional scaling at times near zero was important. To quantify the perturbation, the authors of Cass and Friz (2010) introduced an easy to verify coupling condition (see Corollary 2 in Cass and Friz 2010). It should be possible to employ a version of this condition in the present paper to justify the replacement mentioned above. We will however not pursue this point further here.

REMARK 2.3 Throughout this article, one can consider a classical (Markovian, diffusion) stochastic volatility setting by taking $K \equiv 1$, or equivalently $H \equiv 1/2$, by simply ignoring all hats ($\hat{\cdot}$) in the sequel. In particular then, $\hat{\varepsilon}/\varepsilon \equiv 1$ in all subsequent formulae.

General facts on large deviations of Gaussian measures on Banach spaces (Deuschel and Stroock 1989) such as the path space $C([0,1],\mathbb{R}^3)$ imply that a large deviation principle holds for the triple $\{\hat{\varepsilon}(W,B,\hat{B}) : \hat{\varepsilon} > 0\}$, with speed $\hat{\varepsilon}^2$ and rate function

$$\begin{cases} \frac{1}{2}\|h\|^2_{H_0^1} + \frac{1}{2}\|f\|^2_{H_0^1}, & f,h \in H_0^1 \text{ and } \hat{f} = K\dot{f}, \\ +\infty, & \text{otherwise,} \end{cases} \tag{4}$$

where

$$K\dot{f}(t) := \int_0^t K(t,s)\dot{f}(s)\,\mathrm{d}s$$

for $f \in H_0^1$, the space of absolutely continuous paths with L^2 derivative

$$H_0^1 := \left\{ f : [0,1] \to \mathbb{R} \text{continuous} \,\middle|\, \|f\|^2_{H_0^1} := \int_0^1 \left|\dot{f}(s)\right|^2 \mathrm{d}s < \infty, f(0) = 0 \right\}. \tag{5}$$

This enables us to derive a large deviations principle for X in (3): the (local) small-time self-similarity property of $\hat{B}$ (Assumption 2.1) implies that $X_t \overset{law}{=} X_1^\varepsilon$ where

$$\mathrm{d}X_t^\varepsilon = \sigma(\hat{\varepsilon}\hat{B}_t)\varepsilon\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) - \tfrac{1}{2}\varepsilon^2\sigma^2(\hat{\varepsilon}\hat{B}_t)\,\mathrm{d}t, \quad X_0^\varepsilon = 0.$$

For what follows, it will be convenient to consider a rescaled version of (3)

$$\mathrm{d}\hat{X}_t^\varepsilon \equiv \mathrm{d}\left(\frac{\hat{\varepsilon}}{\varepsilon}X_t^\varepsilon\right) = \sigma(\hat{\varepsilon}\hat{B}_t)\hat{\varepsilon}\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) - \frac{1}{2}\varepsilon\hat{\varepsilon}\sigma^2(\hat{\varepsilon}\hat{B}_t)\,\mathrm{d}t, \quad \hat{X}_0^\varepsilon = 0.$$

Under a linear growth condition on the function σ, Forde and Zhang (2017) use the extended contraction principle to establish a large deviations principle for $(\hat{X}_1^\varepsilon)$ with speed $\hat{\varepsilon}^2$. More precisely, with

$$\varphi_1(h,f) := \Phi_1(h,f,\hat{f}) = \int_0^1 \sigma(\hat{f})\,\mathrm{d}(\bar{\rho}h + \rho f), \tag{6}$$

the rate function is given by

$$\begin{aligned} I(x) &= \inf_{h,f\in H_0^1}\left\{\frac{1}{2}\int_0^1 \dot{h}^2\,\mathrm{d}t + \frac{1}{2}\int_0^1 \dot{f}^2\,\mathrm{d}t : \varphi_1(h,f) = x\right\} \\ &= \inf_{f\in H_0^1}\left\{\frac{1}{2}\frac{\left(x - \rho\left\langle\sigma(\hat{f}),\dot{f}\right\rangle\right)^2}{\bar{\rho}^2\left\langle\sigma^2(\hat{f}),1\right\rangle} + \frac{1}{2}\int_0^1 \dot{f}^2\,\mathrm{d}t\right\}, \end{aligned} \tag{7}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product on $L^2([0,1],dt)$. Several other proofs (under varying assumptions on σ) have appeared since (Jacquier *et al.* 2017; Bayer *et al.* 2017; Gulisashvili 2017).

As a matter of fact, this paper relies on moderate – rather than large – deviations, as emphasized in (iiic) below. To this end, let us make

ASSUMPTION 2.4

(i) (*Positive spot vol*) *Assume* $\sigma : \mathbb{R} \to \mathbb{R}$ *is smooth with* $\sigma_0 := \sigma(0) > 0$.
(ii) (*Roughness*) *The Hurst parameter* H *satisfies* $H \in (0, 1/2]$.
(iiia) (*Martingality*) *The price process* $S = \exp X$ *is a martingale.*
(iiib) (*Short-time moments*) $\forall m < \infty\ \exists t > 0 : E(S_t^m) < \infty$.

While condition (iiia) hardly needs justification, we emphasize that conditions (iiia-b) are only used to the extent that they imply condition (iiic) given below (which thus may replace (iiia-b) as an alternative, if more technical, assumption). The reason we point this out explicitly is that all the conditions (iiia-c) are implicit (growth) conditions on the function $\sigma(.)$. For instance, (iiia-b) was seen to hold under a linear growth assumption (Forde and Zhang 2017; Gulisashvili 2017), whereas the log-normal volatility case (think of $\sigma(x) = \mathrm{e}^x$) is complicated. Martingality, for instance, requires $\rho \le 0$ and there is a critical moment $m^* = m^*(\rho)$, even when $\rho < 0$. See Sin (1998), Jourdain (2004) and Lions and Musiela (2007) for the case $H = 1/2$ and the forthcoming work (Friz and Gassiat 2018) for the general rough case $H \in (0,1]$. We view (iiic) simply as a more flexible condition that can hold in situations where (iiib) fails.

(iiic) (Call price upper moderate deviation bound) For every $\beta \in (0,H)$, and every fixed $x>0$, and $\hat{x}_\varepsilon := x\varepsilon^{1-2H+2\beta}$,

$$E[(\mathrm{e}^{X_1^\varepsilon} - \mathrm{e}^{\hat{x}_\varepsilon})^+] \le \exp\left(-\frac{x^2+o(1)}{2\sigma_0^2\varepsilon^{4H-4\beta}}\right).$$

This condition is reminiscent of the 'upper part' of the large deviation estimate obtained in Forde and Zhang (2017)

$$E[(\mathrm{e}^{X_1^\varepsilon} - \mathrm{e}^{x\varepsilon^{1-2H}})^+] = \exp\left(-\frac{I(x)+o(1)}{\varepsilon^{4H}}\right). \tag{8}$$

If fact, if one *formally* applies this with x replaced by $x\varepsilon^{2\beta}$, followed by Taylor expanding the rate function,

$$I(x\varepsilon^{2\beta}) \sim \frac{1}{2}I''(0)x^2\varepsilon^{4\beta} = \frac{1}{2\sigma_0^2}x^2\varepsilon^{4\beta},$$

one readily arrives at the estimate (iiic). Unfortunately, $o(1) = o_x(1)$ in (8), which is a serious obstacle in making this argument rigorous. Instead, we will give a direct argument (Lemma 7.1) to see how (iiia-b) implies (iiic).

In the sequel, we will use another mild assumption on the kernel.

Assumption 2.5 *The kernel K has the following properties*

(i) $\hat{B}_t = \int_0^t K(t,s)\,\mathrm{d}B_s$ *has a continuous (in t) version on* $[0,1]$.
(ii) $\forall t \in [0,1] : \int_0^t K(t,s)^2\,\mathrm{d}s < \infty$.

Note that the Riemann-Liouville kernel $K(t,s) = \sqrt{2H}(t-s)^\gamma$, $\gamma = H - 1/2$ satisfies Assumption 2.5.

Remark 2.6 Assumption 2.5 implies that the Cameron-Martin space $\mathcal{H}$ of $\hat{B}$ is given by the image of H_0^1 under K, i.e.

$$\mathcal{H} = \{K\dot{f} \mid f \in H_0^1\}.$$

See Lemma 5.3 and Remark 5.4 for more details. A reference and also a sufficient condition for Assumption 2.5 (i) can be found e.g. in Decreusefond (2005, Section 3).

3. Main results

The following result can be seen as a non-Markovian extension of work by Osajima (2015). The statement here is a combination of Theorem 5.10 and Proposition (5.14) below. Recall that $\sigma_0 = \sigma(0)$ represents spot-volatility. We also set $\sigma_0' \equiv \sigma'(0)$.

Theorem 3.1 Energy expansion *The rate function (or energy) I in (7) is smooth in a neighborhood of $x=0$ (at-the-money) and it is of the form*

$$I(x) = \frac{1}{\sigma_0^2}\frac{x^2}{2} - \left(6\rho\frac{\sigma_0'}{\sigma_0^4}\int_0^1\int_0^t K(t,s)\,\mathrm{d}s\,\mathrm{d}t\right)\frac{x^3}{3!} + \mathcal{O}(x^4).$$

The next result is an exact representation of call prices, valid in a non-Markovian generality, and amenable to moderate- and large-deviation analysis (Theorem 3.4 below).

Theorem 3.2 Pricing formula *For a fixed log-strike $x \ge 0$ and time to maturity $t>0$, set $\hat{x} := (\varepsilon/\hat{\varepsilon})x$, where $\varepsilon = t^{1/2}$ and $\hat{\varepsilon} = t^H = \varepsilon^{2H}$, as before. Then we have*

$$\begin{aligned} c(\hat{x},t) &= E\left[\left(\exp(X_t) - \exp\hat{x}\right)^+\right] \\ &= \mathrm{e}^{-I(x)/\hat{\varepsilon}^2}\,\mathrm{e}^{\varepsilon/\hat{\varepsilon}x}J(\varepsilon,x), \end{aligned} \tag{9}$$

where

$$J(\varepsilon,x) := E\left[\mathrm{e}^{-(I'(x)/\hat{\varepsilon}^2)\hat{U}^\varepsilon}\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{U}^\varepsilon\right) - 1\right)\mathrm{e}^{I'(x)R_2^\varepsilon}\mathbf{1}_{\hat{U}^\varepsilon \ge 0}\right]$$

and $\hat{U}^\varepsilon$ is a random variable of the form

$$\hat{U}^\varepsilon = \hat{\varepsilon}g_1 + \hat{\varepsilon}^2R_2^\varepsilon \tag{10}$$

with g_1 a centred Gaussian random variable, explicitly given in equation (38) below, and R_2^ε is a (random) remainder term, in the sense of a stochastic Taylor expansion in $\hat{\varepsilon}$, see Lemma 6.2 for more details.

Example 3.3 Black-Scholes model We fix volatility $\sigma(\cdot) \equiv \sigma > 0$, and $H = 1/2$ so that $\hat{\varepsilon} = \varepsilon$ and all $\hat{\cdot}$ can be omitted. Energy is given by $I(x) = x^2/2\sigma^2$ and

$$U^\varepsilon = \varepsilon g_1 + \varepsilon^2 R_2^\varepsilon \equiv \varepsilon\sigma W_1 - \varepsilon^2\sigma^2/2$$

with $R_2^\varepsilon = R_2 \equiv -\sigma^2/2$ independent of ε. Moreover,

$$\begin{aligned} J(\varepsilon,x) &= E\left[\mathrm{e}^{-(I'(x)/\varepsilon^2)U^\varepsilon}\left(\mathrm{e}^{U^\varepsilon} - 1\right)\mathrm{e}^{I'(x)R_2}\mathbf{1}_{U^\varepsilon \ge 0}\right] \\ &= E\left[\mathrm{e}^{-(I'(x)/\varepsilon)g_1}\left(\mathrm{e}^{\varepsilon g_1 - \varepsilon^2\sigma^2/2} - 1\right)\mathbf{1}_{\{g_1 \ge \varepsilon\sigma^2/2\}}\right] \\ &= E\left[\mathrm{e}^{-\alpha W_1}\left(\mathrm{e}^{\varepsilon\sigma W_1 - (\varepsilon\sigma)^2/2} - 1\right)\mathbf{1}_{\{W_1 \ge \varepsilon\sigma/2\}}\right] \\ &= \mathrm{e}^{-(\varepsilon\sigma)^2/2}M(-\alpha + \varepsilon\sigma) - M(-\alpha) \end{aligned} \tag{11}$$

with $\alpha := I'(x)\sigma/\varepsilon = (1/\sigma)(x/\varepsilon)$, and, in terms of the standard Gaussian cdf Φ,

$$M(\beta) := E\left[\mathrm{e}^{\beta W_1}\mathbf{1}_{\{W_1 \ge \varepsilon\sigma/2\}}\right] = \mathrm{e}^{\beta^2/2}\Phi\left(\beta - \tfrac{\varepsilon\sigma}{2}\right).$$

Using the expansion $\Phi(-y) = (1/y\sqrt{2\pi})\,\mathrm{e}^{-y^2/2}(1 - y^{-2} + \cdots)$, as $y \to \infty$ one deduces, for fixed $x>0$, the asymptotic relation, as $\varepsilon \to 0$,

$$J(\varepsilon,x) \sim \frac{\mathrm{e}^{-x/2}}{\sqrt{2\pi}}\frac{\varepsilon^3\sigma^3}{x^2}. \tag{12}$$

We will be interested (cf. Theorem 3.4) in replacing x by $\tilde{x} = x\varepsilon^{2\beta} \to 0$ for $\beta > 0$. This gives $\tilde{\alpha} = (1/\sigma)(x/\varepsilon^{1-2\beta})$ and the above analysis, now based on $\tilde{\alpha} \to \infty$, remains valid† for β in the 'moderate' regime $\beta \in [0,1/2)$ and we obtain

$$\forall\, x>0, \quad \beta \in [0,1/2) : J\left(\varepsilon, x\varepsilon^{2\beta}\right) \sim \frac{1}{\sqrt{2\pi}}\frac{\varepsilon^{3-4\beta}\sigma^3}{x^2}. \tag{13}$$

Let us point out, for the sake of completeness, that a similar expansion is *not* valid for $\beta > 1/2$. To see this, first note

† More terms in the expansion of Φ are needed.

that (9) implies that $J(\varepsilon,x)|_{x=0}$ is precisely the ATM call price with time $t=\varepsilon^2$ from expiration. Well-known ATM asymptotics then imply that $J(\varepsilon,x)|_{x=0} \sim (1/\sqrt{2\pi})\varepsilon\sigma$ as $\varepsilon \to 0$. These asymptotics are unchanged in case of $o(t^{1/2}) = o(\varepsilon)$ out-of-moneyness ('almost-at-the-money' in the terminology of Friz *et al*. 2018), which readily implies

$$\forall\, x>0, \quad \beta > 1/2 : J\left(\varepsilon, x\varepsilon^{2\beta}\right) \sim \frac{1}{\sqrt{2\pi}}\varepsilon\sigma = \text{const} \times \varepsilon$$

At last, we have the borderline case $\beta = 1/2$, or $\tilde{x} = x\varepsilon$. From e.g. Muhle-Karbe and Nutz (2011, Theorem 3.1), we see that $c(x\varepsilon, \varepsilon^2) \sim a(x;\sigma)\varepsilon$ with positive constant $a(x;\sigma)$. A look at (9) then reveals

$$\forall\, x>0 : J\left(\varepsilon, x\varepsilon\right) \sim a(x;\sigma)\varepsilon\, \mathrm{e}^{x^2/2\sigma^2} = \text{const} \times \varepsilon.$$

For the call price expansion in the large / moderate deviations regime, $\beta \in [0, 1/2)$, the polynomial in ε-behavior of (13) implies that the J-term in the pricing formula will be negligible on the moderate / large deviation scale, in the sense for any $\theta > 0$, we have $\varepsilon^\theta \log J(\varepsilon, x\varepsilon^{2\beta}) \to 0$ as $\varepsilon \to 0$. Consequently, with $k_t = kt^\beta$, for $t=\varepsilon^2$, $k>0$, $\beta \in [0,1/2)$, we get the 'moderate' Black-Scholes call price expansion,

$$-\log c_{BS}(k_t, t) = \frac{1}{t^{1-2\beta}}\frac{k^2}{2\sigma^2}\left(1 + o\left(1\right)\right) \quad \text{as } t \downarrow 0.$$

While the above can be confirmed by elementary analysis of the Black–Scholes formula, the following theorem exhibits it as an instance of a general principle. See Friz *et al*. (2018) for a general diffusion statement.

THEOREM 3.4 Moderate Deviations *In the rough volatility regime $H \in (0, 1/2]$, consider log-strikes of the form*

$$k_t = kt^{1/2-H+\beta} \quad \text{for a constant } \ k \geq 0.$$

(i) *For $\beta \in (0, H)$, and every $\theta > 0$, we have*

$$-\log c(k_t, t) = \frac{I''(0)}{t^{2H-2\beta}}\frac{k^2}{2} + O(t^{3\beta-2H}) + O(t^{-\theta}) \quad \text{as } t \downarrow 0.$$

(ii) *For $\beta \in (0, \frac{2}{3}H)$, and every $\theta > 0$, we have*

$$-\log c(k_t, t) = \frac{I''(0)}{t^{2H-2\beta}}\frac{k^2}{2} + \frac{I'''(0)}{t^{2H-3\beta}}\frac{k^3}{6} + O(t^{4\beta-2H}) + O(t^{-\theta}) \quad \text{as} t \downarrow 0.$$

Moreover,

$$I''(0) = \frac{1}{\sigma_0^2},$$

$$I'''(0) = -6\rho\frac{\sigma_0'}{\sigma_0^4}\int_0^1\int_0^t K(t,s)\,\mathrm{d}s\,\mathrm{d}t = -6\rho\frac{\sigma_0'}{\sigma_0^4}\langle K1, 1\rangle,$$

where $\langle\cdot,\cdot\rangle$ is the inner product in $L^2([0,1])$.

REMARK 3.5 In principle, further terms (of order $t^{i\beta-2H}$, $i=4,5,\ldots$) can be added to this expansion of log call prices, given that the energy has sufficient regularity, see Theorem 3.6. We also note that, for small enough β, the error term $O(t^{-\theta})$ can be omitted. In any case, one can replace the additive error bounds by (cruder) ones, where the right-most term in the expansion is multiplied with $(1+o(1))$, as was done in Friz *et al*. (2018).

Proof of Theorem 3.4 We apply Theorem 3.2 with $\hat{x} = k_t = kt^{1/2-H+\beta}$, i.e. with $x = kt^\beta = k\varepsilon^{2\beta}$. In particular, we so get, with $\hat{\varepsilon} = t^H$ and $\varepsilon = t^{1/2}$,

$$c(k_t, t) = \mathrm{e}^{-I(x)/\hat{\varepsilon}^2}\,\mathrm{e}^{\varepsilon/\hat{\varepsilon}x}\,J\left(\varepsilon, k\varepsilon^{2\beta}\right).$$

The technical Proposition 7.3 asserts that, for fixed $k>0$, the factor J is negligible in the sense that, for every $\theta > 0$,

$$\varepsilon^\theta \log J(\varepsilon, k\varepsilon^{2\beta}) \to 0 \quad \text{as } \varepsilon \to 0.$$

The theorem now follows immediately from the Taylor expansion of $I(x)$ around $x=0$ (see Theorem 3.1), plugging in $x = kt^\beta$. Indeed, replacing $I(x)$ by the Taylor-jet seen in (i),(ii), leads exactly to an error term $O(t^{3\beta-2H})$, resp. $O(t^{4\beta-2H})$. ■

Fix real numbers $k>0$, $0<H<\frac{1}{2}$, $0<\beta<H$, and an integer $n \geq 2$. For every $t>0$, set

$$k_t = kt^{1/2-H+\beta},$$

and denote

$$\phi_{n,H,\beta,\theta}(t) = \max\left\{t^{2H-2\beta-\theta},\ t^{(n-1)\beta}\right\}.$$

Here, $\theta > 0$ can be arbitrarily small. It is clear that for all small t and θ small enough,

$$\phi_{n,H,\beta,\theta}(t) = t^{2H-2\beta-\theta} \Leftrightarrow 2H-2\beta \leq (n-1)\beta \Leftrightarrow \frac{2H}{n+1} \leq \beta,$$

while

$$\phi_{n,H,\beta,\theta}(t) = t^{(n-1)\beta} \Leftrightarrow 2H-2\beta > (n-1)\beta \Leftrightarrow \beta < \frac{2H}{n+1}.$$

The following statement provides an asymptotic formula for the implied variance.

THEOREM 3.6 *Suppose $0<\beta<2H/n$ and $\theta>0$ small enough. Then as $t \to 0$ (and for $k>0$),*

$$\sigma_{\text{impl}}(k_t, t)^2 = \sum_{j=0}^{n-2}\frac{(-1)^j 2^j}{I''(0)^{j+1}}\left(\sum_{i=3}^{n}\frac{I^{(i)}(0)}{i!}k^{i-2}t^{(i-2)\beta}\right)^j + \mathcal{O}\left(\phi_{n,H,\beta,\theta}(t)\right). \tag{14}$$

The $\mathcal{O}$-estimate in (14) depends on n, H, β, θ, and k. It is uniform on compact subsets of $[0,\infty)$ with respect to the variable k.

REMARK 3.7 Using the multinomial formula, we can represent the expression on the left-hand side of (14) in terms of

certain powers of t. However, the coefficients become rather complicated.

REMARK 3.8 Let an integer $n \geq 2$ be fixed, and suppose we would like to use only the derivatives $I^{(i)}(0)$ for $2 \leq i \leq n$ in formula (14) to approximate $\sigma_{\text{impl}}(k_t, t)^2$. Then, the optimal range for β is the following: $2H/(n+1) \leq \beta < 2H/n$. On the other hand, if β is outside of the interval $[2H/(n+1), 2H/n)$, more derivatives of the energy function at zero may be needed to get a good approximation of the implied variance in formula (14).

We will next derive from Theorem 3.6 several asymptotic formulas for the implied volatility. In the next corollary, we take $n=2$.

COROLLARY 3.9 *As* $t \to 0$,

$$\sigma_{\text{impl}}(k_t, t) = \sigma_0 + \mathcal{O}(\phi_{2,H,\beta,\theta}(t)). \tag{15}$$

Corollary 3.9 follows from Theorem 3.6 with $n=2$, the equality

$$I''(0) = \sigma_0^{-2} \tag{16}$$

given in Theorem 3.4, and the Taylor expansion $\sqrt{1+h} = 1 + \mathcal{O}(h)$ as $h \to 0$.

In the next corollary, we consider the case where $n=3$.

COROLLARY 3.10 *Suppose* $\beta < 2H/3$. *Then, as* $t \to 0$,

$$\sigma_{\text{impl}}(k_t, t) = \sigma_0 + \rho \frac{\sigma_0'}{\sigma_0} \langle K1, 1\rangle k t^{\beta} + \mathcal{O}(\phi_{3,H,\beta,\theta}(t)). \tag{17}$$

Corollary 3.10 follows from Theorem 3.6 with $n=3$, formula (16), the equality

$$I'''(0) = -6\rho \frac{\sigma_0'}{\sigma_0^4} \langle K1, 1\rangle \tag{18}$$

(see Theorem 3.4), and the expansion $\sqrt{1+h} = 1 + \frac{1}{2}h + \mathcal{O}(h^2)$ as $h \to 0$.

Using Corollary 3.10, we establish the following implied volatility skew formula in the moderate deviation regime.

COROLLARY 3.11 *Let* $0 < H < \frac{1}{2}$, $0 < \beta < \frac{2}{3}H$, *and fix* $y,z > 0$ *with* $y \neq z$. *Then as* $t \to 0$,

$$\frac{\sigma_{\text{impl}}(yt^{1/2-H+\beta}, t) - \sigma_{\text{impl}}(zt^{1/2-H+\beta}, t)}{(y-z)t^{1/2-H+\beta}} \sim \rho \frac{\sigma_0'}{\sigma_0} \langle K1, 1\rangle t^{H-1/2}. \tag{19}$$

REMARK 3.12 Corollary 3.11 complements earlier works of Alòs *et al.* (2007) and Fukasawa (2011, 2017). For instance, the following formula can be found in Fukasawa (2017, p. 6), see also Fukasawa (2011, p. 14):

$$\frac{\sigma_{\text{impl}}(yt^{1/2}, t) - \sigma_{\text{impl}}(zt^{1/2}, t)}{(y-z)t^{1/2}} \sim \rho C(H) \frac{\sigma_0'}{\sigma_0} t^{H-1/2}. \tag{20}$$

In formula (20), we employ the notation used in the present paper. Our analysis shows that the applicability range of skew approximation formulas is by no means restricted to the Central Limit Theorem type log-moneyness deviations of order $t^{1/2}$. It also includes the moderate deviations regime of order $t^{1/2-H+\beta}$. The previous rate is clearly $\gg t^{1/2}$ as $t \to 0$.

REMARK 3.13 Symmetry Write $\Phi_1(W, B, \hat{B}; \rho; \sigma)$ for the 'Itô-type map'

$$\Phi_1(W, B, \hat{B}) := \int_0^1 \sigma(\hat{B})\, \mathrm{d}\,(\bar{\rho} W + \rho B)\,.$$

It equals, in law, $\Phi_1(W, -B, -\hat{B}; -\rho; \sigma(-\cdot))$, and indeed all our formulae are invariant under this transformation. In particular, the skew remains unchanged when the pair (ρ, σ_0') is replaced by $(-\rho, -\sigma_0')$.

4. Simulation results

We verify our theoretical results numerically with a variant of the *rough Bergomi model* (Bayer *et al.* 2016) which fits nicely into the general rough volatility framework considered in this paper. As before, the model has been normalized such that $S_0 = 1$ and $r=0$. We let (W, B) be two independent Brownian motions and $\rho \in (-1, 1)$ with $\bar{\rho}^2 = 1 - \rho^2$ such that $Z = \bar{\rho} W + \rho B$ is another Brownian motion having constant correlation ρ with B. For some spot volatility σ_0 and volatility of volatility parameter η, we then assume the following dynamics for some asset S:

$$\frac{dS_t}{S_t} = \sigma(\hat{B}_t) dZ_t \tag{21}$$

$$\sigma(x) = \sigma_0 \exp\left(\frac{1}{2}\eta x\right) \tag{22}$$

where $\hat{B}$ is a Riemann-Liouville fBM given by

$$\hat{B}_t = \sqrt{2H} \int_0^t |t-s|^{H-1/2}\, \mathrm{d}B_s.$$

The approach taken for the Monte Carlo simulations of the quantities we are interested in is the one initially explored in the original *rough Bergomi* pricing paper (Bayer *et al.* 2016). That is, exploiting their joint Gaussianity, where we use the well-known Cholesky method to simulate the joint paths of $(Z, \hat{B})$ on some discretization grid $\mathcal{D}$. With (22) being an explicit function in terms of the rough driver, an Euler discretisation of the Ito SDE (21) on $\mathcal{D}$ then yields estimates for the price paths.

The Cholesky algorithm critically hinges on the availability and explicit computability of the joint covariance matrix of $(Z, \hat{B})$ whose terms we readily compute below.†

LEMMA 4.1 *For convenience, define constants* $\gamma = \frac{1}{2} - H \in [0, \frac{1}{2})$ *and* $D_H = \sqrt{2H}/(H + \frac{1}{2})$ *and define an auxiliary*

† Note that expressions for the exact same scenario have have been computed before in the original pricing paper (Bayer *et al.* 2016), yet in that version the expression for the autocorrelation of the fBM $\hat{B}$ was incorrect. We compute and state here all the relevant terms for the sake of completeness.

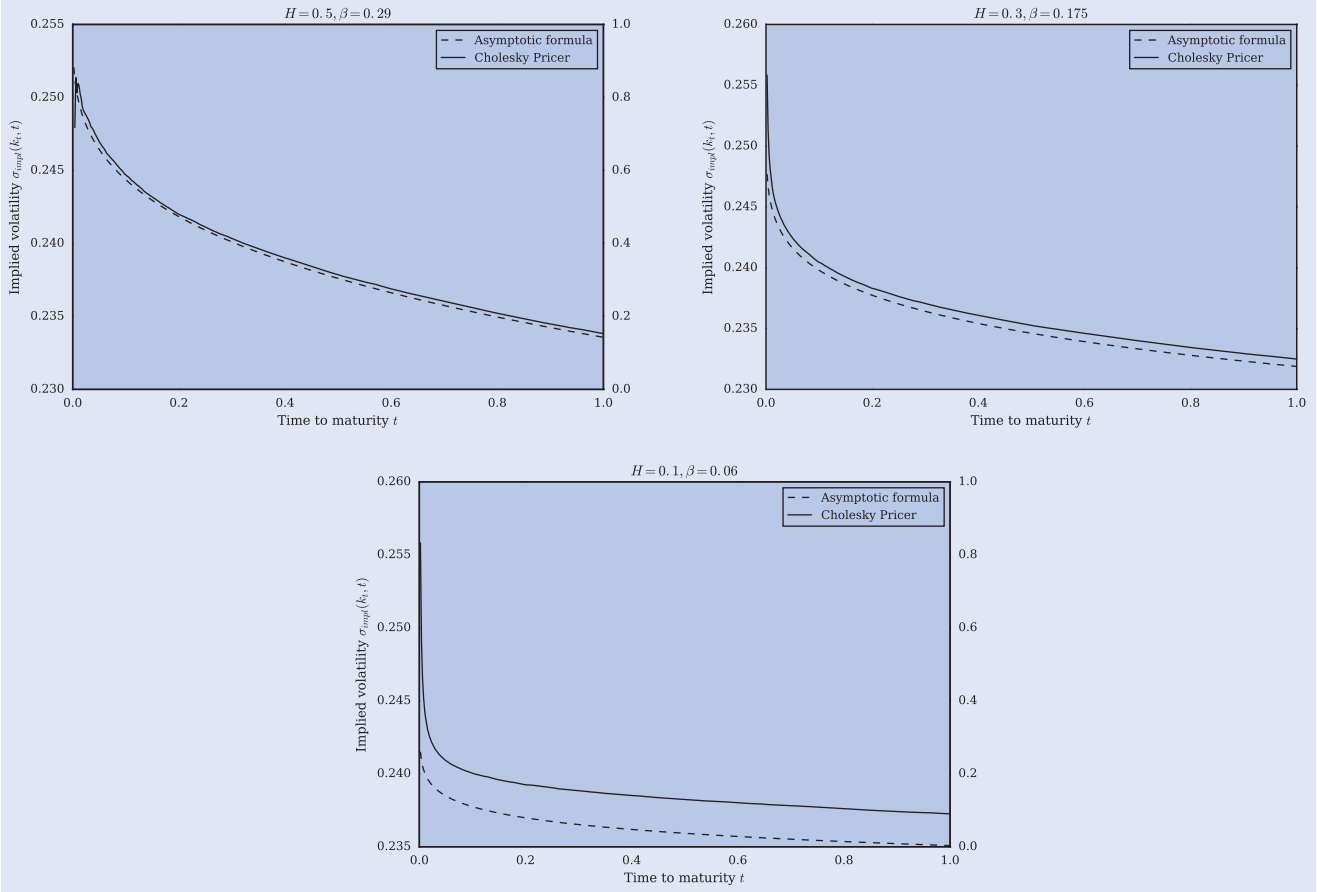

Figure 1. Illustration of the term structure of implied volatility of the Modified Rough Bergomi model in the Moderate deviations regime with time-varying log-strike $k_t = 0.4t^\beta$. Depicted are the asymptotic formula (equation (17), dashed line) and an estimate based on $N = 10^8$ samples of a MC Cholesky Option Pricer (solid line) with 500 time steps. Model parameters are given by spot vol $\sigma_0 \approx 0.2557$, vvol $\eta = 0.2928$ and correlation parameter $\rho = -0.7571$.

function $G : [1, \infty) \to \mathbb{R}$ *by*

$$G(x) = 2H\left(\frac{1}{1-\gamma}x^{-\gamma} + \frac{\gamma}{1-\gamma}x^{-(1+\gamma)} \times \frac{1}{2-\gamma}{}_2F_1(1, 1+\gamma, 3-\gamma, x^{-1})\right) \tag{23}$$

where ${}_2F_1$ *denotes the Gaussian hypergeometric function* (Olver *et al.* 2010). *Then the joint process* $(Z, \hat{B})$ *has zero mean and covariance structure governed by*

$$\begin{aligned}
&\mathrm{Var}[\hat{B}_t^2] = t^{2H}, && \text{for } t \geq 0,\\
&\mathrm{Cov}[\hat{B}_s\hat{B}_t] = t^{2H}G(s/t), && \text{for } s > t \geq 0,\\
&\mathrm{Cov}[\hat{B}_sZ_t] = \rho D_H(s^{H+1/2}\\
&-(s - \min(t,s))^{H+1/2}), && \text{for } t, s \geq 0,\\
&\mathrm{Cov}[Z_tZ_s] = \min(t,s), && \text{for } t, s \geq 0.
\end{aligned}$$

Numerical simulations† confirm the theoretical results obtained in the last section. In particular – as can be seen in figure 1 – the asymptotic formula for the implied volatility (17) captures very well the geometry of the term structure of implied volatility, with particularly good results for higher H and worsening results as $H \downarrow 0$. Quite surprisingly, despite being an asymptotic formula, it seems to be fairly accurate over a wide array of maturities extending up to a single year.

† The Python 3 code used to run the simulations can be found at github.com/RoughStochVol.

5. Proof of the energy expansion

Consider

$$\begin{aligned}
\mathrm{d}X &= -\tfrac{1}{2}\sigma^2(Y)\,\mathrm{d}t + \sigma\,(Y)\,\mathrm{d}\,(\bar{\rho}\,\mathrm{d}W + \rho\,\mathrm{d}B)\,, \quad X_0 = 0\\
\mathrm{d}Y &= \mathrm{d}\hat{B}, \quad Y_0 = 0
\end{aligned}$$

where $\hat{B}_t = \int_0^t K(t,s)\,\mathrm{d}B_s$ for a fixed Volterra kernel (recall (3) in the previous section). We study the small noise problem $(X^\varepsilon, Y^\varepsilon)$ where $(W, B, \hat{B})$ is replaced by $(\varepsilon W, \varepsilon B, \hat{\varepsilon}\hat{B})$. The following proposition roughly says that

$$\mathbb{P}\left(X_1^\varepsilon \approx \frac{\varepsilon}{\hat{\varepsilon}}x\right) \approx \exp\left(-\frac{I(x)}{\hat{\varepsilon}^2}\right).$$

Proposition 5.1 Forde and Zhang 2017 *Under suitable assumptions* (*cf. Section* 2), *the rescaled process* $((\hat{\varepsilon}/\varepsilon)X_1^\varepsilon$:

$\varepsilon \geq 0$) *satisfies an LDP* (*with speed* $\hat{\varepsilon}^2$) *and rate function*

$$I(x) = \inf_{f \in H_0^1} \left[\frac{(x - \rho G(f))^2}{2\bar{\rho}^2 F(\hat{f})} + \frac{1}{2} E(f) \right] \equiv \inf_{f \in H_0^1} \mathcal{I}_x(f) \equiv \mathcal{I}_x(f^x), \tag{24}$$

where

$$G(f) = \int_0^1 \sigma\left(\left(K\dot{f}\right)(s)\right) \dot{f}_s \, \mathrm{d}s \equiv \left\langle \sigma\left(K\dot{f}\right), \dot{f} \right\rangle \equiv \left\langle \sigma(\hat{f}), \dot{f} \right\rangle$$

$$F(f) = \int_0^1 \sigma\left(\left(K\dot{f}\right)(s)\right)^2 \mathrm{d}s \equiv \left\langle \sigma^2\left(K\dot{f}\right), 1 \right\rangle \equiv \left\langle \sigma^2(\hat{f}), 1 \right\rangle$$

$$E(f) = \int_0^1 \left|\dot{f}(s)\right|^2 \mathrm{d}s \equiv \left\langle \dot{f}, \dot{f} \right\rangle$$

The rest of this section is devoted to analysis of the function I as defined in (24). First, we derive the first order optimality condition for the above minimization problem.

Proposition 5.2 First order optimality condition *For any* $x \in \mathbb{R}$ *we have at any local minimizer* $f = f^x$ *of the functional* $\mathcal{I}_x$ *in* (24) *that*

$$f_t^x = \frac{\rho\left(x - \rho G(f^x)\right)\left\{\left\langle \sigma\left(K\dot{f}^x\right), 1_{[0,t]} \right\rangle + \left\langle \sigma'\left(K\dot{f}^x\right)\dot{f}^x, K1_{[0,t]} \right\rangle\right\}}{\bar{\rho}^2 F(f^x)} + \frac{(x - \rho G(f^x))^2}{\bar{\rho}^2 F^2(f^x)} \left\langle \left(\sigma\sigma'\right)\left(K\dot{f}^x\right), K1_{[0,t]} \right\rangle, \tag{25}$$

for all $t \in [0,1]$.

Proof We denote $a \approx b$ whenever $a = b + o(\delta)$ for a small parameter δ. We expand

$$E(f + \delta g) \approx E(f) + 2\delta \left\langle \dot{f}, \dot{g} \right\rangle$$

$$F(f + \delta g) \approx F(f) + \delta \left\langle \left(\sigma^2\right)'\left(K\dot{f}\right), K\dot{g} \right\rangle$$

$$G(f + \delta g) \approx G(f) + \delta \left\{ \left\langle \sigma\left(K\dot{f}\right), \dot{g} \right\rangle + \left\langle \sigma'\left(K\dot{f}\right)\dot{f}, K\dot{g} \right\rangle \right\}$$

If $f = f^x$ is a minimizer then $\delta \mapsto \mathcal{I}_x(f + \delta g)$ has a minimum at $\delta = 0$ for all g. We expand

$$\begin{aligned}
\mathcal{I}_x(f + \delta g) &= \frac{(x - \rho G(f + \delta g))^2}{2\bar{\rho}^2 F(f + \delta g)} + \frac{1}{2} E(f + \delta g) \\
&\approx \frac{\left(x - \rho G(f) - \delta\rho\left\{\left\langle \sigma\left(K\dot{f}\right), \dot{g} \right\rangle + \left\langle \sigma'\left(K\dot{f}\right)\dot{f}, K\dot{g} \right\rangle\right\}\right)^2}{2\bar{\rho}^2\left[F(f) + \delta\left\langle \left(\sigma^2\right)'\left(K\dot{f}\right), K\dot{g} \right\rangle\right]} \\
&\quad + \frac{1}{2} E(f) + \delta\left\langle \dot{f}, \dot{g} \right\rangle \\
&\approx \frac{(x - \rho G(f))^2 - \delta 2\rho\,(x - \rho G(f))\left\{\left\langle \sigma\left(K\dot{f}\right), \dot{g} \right\rangle + \left\langle \sigma'\left(K\dot{f}\right)\dot{f}, K\dot{g} \right\rangle\right\}}{2\bar{\rho}^2 F(f)\left[1 + \frac{\delta}{F(f)}\left\langle \left(\sigma^2\right)'\left(K\dot{f}\right), K\dot{g} \right\rangle\right]} \\
&\quad + \frac{1}{2} E(f) + \delta\left\langle \dot{f}, \dot{g} \right\rangle \\
&\approx \frac{(x - \rho G(f))^2 - \delta 2\rho\,(x - \rho G(f))}{2\bar{\rho}^2 F(f)} \left\{\left\langle \sigma\left(K\dot{f}\right), \dot{g} \right\rangle + \left\langle \sigma'\left(K\dot{f}\right)\dot{f}, K\dot{g} \right\rangle\right\} \\
&\quad - \frac{(x - \rho G(f))^2}{2\bar{\rho}^2 F(f)} \frac{\delta}{F(f)} \left\langle \left(\sigma^2\right)'\left(K\dot{f}\right), K\dot{g} \right\rangle + \frac{1}{2} E(f) \\
&\quad + \delta\left\langle \dot{f}, \dot{g} \right\rangle.
\end{aligned}$$

As a consequence, we must have, for $f = f^x$ and every $\dot{g} \in L^2[0,1]$

$$\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}\delta}\left\{\mathcal{I}_x(f + \delta g)\right\}_{\delta=0} \\
&= -\frac{\rho\left(x - \rho G(f)\right)\left\{\left\langle \sigma\left(K\dot{f}\right), \dot{g} \right\rangle + \left\langle \sigma'\left(K\dot{f}\right)\dot{f}, K\dot{g} \right\rangle\right\}}{\bar{\rho}^2 F(f)} \\
&\quad - \frac{(x - \rho G(f))^2}{\bar{\rho}^2 F^2(f)} \left\langle \left(\sigma\sigma'\right)\left(K\dot{f}\right), K\dot{g} \right\rangle + \left\langle \dot{f}, \dot{g} \right\rangle.
\end{aligned}$$

Recall $f_0^x = 0$, any x. We now test with $\dot{g} = 1_{[0,t]}$ for a fixed $t \in [0,1]$ and obtain

$$f_t^x = \frac{\rho\left(x - \rho G(f^x)\right)\left\{\left\langle \sigma\left(K\dot{f}^x\right), 1_{[0,t]} \right\rangle + \left\langle \sigma'\left(K\dot{f}^x\right)\dot{f}^x, K\mathbf{1}_{[0,t]} \right\rangle\right\}}{\bar{\rho}^2 F(f^x)} + \frac{(x - \rho G(f^x))^2}{\bar{\rho}^2 F^2(f^x)} \left\langle \left(\sigma\sigma'\right)\left(K\dot{f}^x\right), K\mathbf{1}_{[0,t]} \right\rangle.$$

■

5.1. Smoothness of the energy

Having formally identified the first order condition for minimality in (24), we will now show that the energy $x \mapsto I(x)$ is a smooth function. More precisely, we will use the implicit function theorem to show that the minimizing configuration f^x is a smooth function in x (locally at $x = 0$). As $\mathcal{I}_x$ is a smooth function, too, this will imply smoothness of $x \mapsto \mathcal{I}_x(f^x) = I(x)$, at least in a neighborhood of 0.

As the Cameron-Martin space $\mathcal{H}$ of the process $\hat{B}$ continuously embeds into $C([0,1])$, K maps H_0^1 continuously into $C([0,1])$, i.e. there is a constant $C > 0$ such that for any $f \in H_0^1$ we have

$$\left\|K\dot{f}\right\|_\infty \leq C \|f\|_{H_0^1}. \tag{26}$$

This result will follow from

Lemma 5.3 *Let* $(V_t : 0 \leq t \leq 1)$ *be a continuous, centred Gaussian process and* $\mathcal{H}$ *its Cameron-Martin space. Then we have the continuous embedding* $\mathcal{H} \hookrightarrow C[0,1]$. *That is, for some constant C,*

$$\|h\|_\infty \leq C \|h\|_{\mathcal{H}}.$$

Proof By a fundamental result of Fernique, applied to the law of V as Gaussian measure on the Banach space

$(C[0,1], \|\cdot\|_\infty)$, the random variable $\|V\|_\infty$ has Gaussian integrability. In particular,

$$\sigma^2 := \mathbb{E}(\|V\|_\infty^2) < \infty,$$

On the other hand, a generic element $h \in \mathcal{H}$ can be written as $h_t = E[V_t Z]$ where Z is a centred Gaussian random variable with variance $\|h\|_{\mathcal{H}}^2$, see, e.g. Friz and Hairer (2014, page 150). By Cauchy–Schwarz,

$$|h_t| \le E\left[|V_t|\right]^{1/2} \|h\|_{\mathcal{H}} \le \sigma \|h\|_{\mathcal{H}}$$

and conclude by taking the sup over on the l.h.s. over $t \in [0,1]$. ■

Remark 5.4 Assume V is of Volterra form, i.e. $V_t = \int_0^t K(t,s)\,\mathrm{d}B_s$. Then it can be shown (see Decreusefond 2005, Section 3) that $\mathcal{H}$ is the image of L^2 under the map

$$K : \dot{f} \mapsto \hat{f} := \left(t \mapsto \int_0^t K(t,s)\dot{f}_s\,\mathrm{d}s\right)$$

and $\|K\dot{f}\|_{\mathcal{H}} = \|\dot{f}\|_{L^2}$. In particular then, applying the above with $h = Kf \in \mathcal{H}$, gives

$$\left\|K\dot{f}\right\|_\infty \le C\left\|K\dot{f}\right\|_{\mathcal{H}} = C\left\|\dot{f}\right\|_{L^2} = C\|f\|_{H_0^1}.$$

5.1.1. The uncorrelated case. We start with the case $\rho = 0$ as the formulas are much simpler in this case.

By Proposition 5.2, any local optimizer $f = f^x$ of the functional $\mathcal{I}_x : H_0^1 \to \mathbb{R}$ in the uncorrelated case $\rho = 0$ satisfies for any $t \in [0,1]$

$$f_t = \frac{x^2}{F^2(f)}\left\langle\left(\sigma\sigma'\right)\left(K\dot{f}\right), K\mathbf{1}_{[0,t]}\right\rangle.$$

We define a map $H : H_0^1 \times \mathbb{R} \to H_0^1$ by

$$H(f,x)(t) := f_t - \frac{x^2}{F^2(f)}\left\langle\left(\sigma\sigma'\right)\left(K\dot{f}\right), K\mathbf{1}_{[0,t]}\right\rangle. \tag{27}$$

Hence, for given $x \in \mathbb{R}$, any local optimizer f must solve $H(f,x) = 0$. As one particular solution is given by the pair $(0,0)$, we are in the realm of the implicit function theorem. We need to prove that

- $(f,x) \mapsto H(f,x)$ is locally smooth (in the sense of Fréchet);
- $DH(f,x) := (\partial/\partial f)H(f,x)$ is invertible in $(0,0)$.

Note that invertibility should hold for x small enough, as $DH(f,x) = \mathrm{id}_{H_0^1} - x^2 R$ for some R, which is invertible as long as R has a bounded norm for sufficiently small x.

Remark 5.5 The method of proof in this section is purely local in H_0^1. Hence, we only really need smoothness of σ locally around 0. Note, however, that stochastic Taylor expansions used in Section 6 will actually require global smoothness of σ.

Lemma 5.6 *The functions* $F : H_0^1 \to \mathbb{R}$ *and* $R_1 : H_0^1 \to C([0,1])$ *defined by*

$$R_1(f)(t) := \left\langle\left(\sigma\sigma'\right)\left(K\dot{f}\right), K\mathbf{1}_{[0,t]}\right\rangle, \quad t \in [0,1],$$

are smooth in the sense of Fréchet.

Proof For $N \ge 1$ we note that the Gateaux derivative of F satisfies

$$D^N F(f)\cdot(g_1,\dots,g_N) = \int_0^1 \frac{\mathrm{d}^N}{\mathrm{d}x^N}\sigma^2(K\dot{f})K\dot{g}_1\cdots K\dot{g}_N\,\mathrm{d}s.$$

By Lemma 5.3, we can bound

$$\begin{aligned}\left|D^N F(f)\cdot(g_1,\dots,g_N)\right| &\le \text{const}\int_0^1 |K\dot{g}_1(s)|\cdots|K\dot{g}_N(s)|\,\mathrm{d}s\\ &\le \text{const}\,\|K\dot{g}_1\|_\infty\cdots\|K\dot{g}_N\|_\infty\\ &\le \text{const}C^N\,\|g_1\|_{H_0^1}\cdots\|g_N\|_{H_0^1},\end{aligned}$$

for $\text{const} = \left\|(\mathrm{d}^n/\mathrm{d}x^n)\sigma^2\right\|_\infty$.† Thus, $D^N F(f)$ is a multilinear form on H_0^1 with operator norm $\left\|D^N F(f)\right\| \le \left\|(\mathrm{d}^n/\mathrm{d}x^n)\sigma^2\right\|_\infty C^N$ independent of f. As $f \mapsto D^N F(f)$ is continuous, we conclude that $D^N F(f)$ as given above is, in fact, a Fréchet derivative.

Let us next consider the functional R_1. Note that

$$\left(D^N R_1(f)\cdot(g_1,\dots,g_N)\right)(t) = \left\langle\mathfrak{s}_N(K\dot{f})K\dot{g}_1\cdots K\dot{g}_N, K\mathbf{1}_{[0,t]}\right\rangle$$

for $\mathfrak{s}_N(x) := (\mathrm{d}^N/\mathrm{d}x^N)\sigma(x)\sigma'(x)$. Hence, Assumption 2.5 implies that

$$\begin{aligned}&\left\|D^N R_1(f)\cdot(g_1,\dots,g_N)\right\|_{H_0^1}^2\\ &= \int_0^1\left(\int_t^1 \mathfrak{s}_N\left((K\dot{f})(s)\right)\prod_{i=1}^N(K\dot{g}_i)(s)K(s,t)\,\mathrm{d}s\right)^2\mathrm{d}t\\ &\le \|\mathfrak{s}_N\|_\infty^2\prod_{i=1}^N\|K\dot{g}_i\|_\infty^2\int_0^1\int_t^1 K(s,t)^2\,\mathrm{d}s\,\mathrm{d}t\\ &\le \|\mathfrak{s}_N\|_\infty^2 C^{2N}\prod_{i=1}^N\|g_i\|_{H_0^1}^2\int_0^1\int_0^s K(s,t)^2\,\mathrm{d}t\,\mathrm{d}s\\ &\le \|\mathfrak{s}_N\|_\infty^2 C^{2N}\int_0^1\int_0^s K(s,t)^2\,\mathrm{d}t\,\mathrm{d}s\prod_{i=1}^N\|g_i\|_{H_0^1}^2.\end{aligned}$$

We see that the multi-linear map $D^N R_1(f)$ has operator norm bounded by

$$\left\|D^N R_1(f)\right\| \le \|\mathfrak{s}_N\|_\infty C^N\sqrt{\int_0^1\int_0^s K(s,t)^2\,\mathrm{d}t\,\mathrm{d}s},$$

independent of f. From continuity of $f \mapsto D^N R_1(f)$, it follows that $D^N R_1(f)$ is the N'th Fréchet derivative. ■

† More precisely, since neither σ nor its derivatives need to be bounded, we need to actually work with a local version of the above estimate, for instance by replacing the max with a sup over a compact set containing $\{(K\dot{f})(t) : 0 \le t \le 1\}$.

THEOREM 5.7 Zero correlation *Assuming* $\rho = 0$, *the energy* $I(x)$ (*as defined in* (24)) *is smooth in a neighborhood of* $x = 0$.

Proof By construction, we have

$$DH(f,x) = \mathrm{id}_{H_0^1} - x^2 A(f)$$

for $A : H_0^1 \to \mathcal{L}(H_0^1, H_0^1)$ defined by

$$A(f) := R_1(f) \otimes DF^{-2}(f) + F^{-2}(f) DR_1(f).$$

Here,

$$\left(R_1(f) \otimes DF^{-2}(f)\right) \cdot g = \underbrace{\left(DF^{-2}(f) \cdot g\right)}_{\in \mathbb{R}} \underbrace{R_1(f)}_{\in H_0^1}.$$

As verified above, H is smooth in the sense of Fréchet. Trivially, $DH(0,0) = \mathrm{id}_{H_0^1}$ is invertible and $H(0,0) = 0$. Therefore, the implicit function theorem implies that there are open neighborhoods U and V of $0 \in H_0^1$ and $0 \in \mathbb{R}$, respectively, and a smooth map $x \mapsto f^x$ from V to U such that $H(f^x, x) \equiv 0$ and f^x is unique in U with this property.

For the energy, we prove that $I(x) = \mathcal{I}_x(f^x)$ in a neighborhood of $x = 0$. First of all, we show that a minimizer exists. If not, there is a function $g \in H_0^1$ with $\mathcal{I}_x(g) < \mathcal{I}_x(f^x)$. For small enough x such a g must be inside a ball with radius ϵ around $0 \in H_0^1$, as $\mathcal{I}_x(g) \geq \frac{1}{2}\|g\|^2_{H_0^1}$ and $\lim_{x\to 0} \mathcal{I}_x(f^x) = 0$. Then note that for any $g \in H_0^1$

$$D^2\mathcal{I}_0(0) \cdot (g,g) = \|g\|^2_{H_0^1} > 0,$$

where $D^2\mathcal{I}_x(f)$ denotes the second derivative of $f \mapsto \mathcal{I}_x(f)$. By continuity, $D^2\mathcal{I}_x(f)$ stays positive definite for (x,f) in a neighborhood of $(0,0)$. As noted, for x small enough, both g and f^x (and the line connecting them) lie in this neighborhood. For $h := g - f^x$, this implies

$$\mathcal{I}_x(g) - \mathcal{I}_x(f^x) = D\mathcal{I}_x(f_x) \cdot h + \int_0^1 D^2\mathcal{I}_x(f^x + th) \cdot (h,h)\, \mathrm{d}t > 0,$$

since $D\mathcal{I}_x(f_x) \cdot h = 0$ and $D^2\mathcal{I}_x(f^x + tsh) \cdot (h,h) > 0$. This contradicts the assumption that $\mathcal{I}_x(g) < \mathcal{I}_x(f^x)$, and we conclude that f^x is, indeed, a minimizer of $\mathcal{I}_x$, implying that $I(x) = \mathcal{I}_x(f^x)$ locally.

Finally, as $x \mapsto f^x$ is smooth and $(f,x) \mapsto \mathcal{I}_x(f) = x^2/2F(f) + \frac{1}{2}\|f\|^2_{H_0^1}$ is smooth, we see that $x \mapsto I(x) = \mathcal{I}_x(f^x)$ is smooth in a neighborhood of 0. (Note that this arguments relies on $\sigma(0) \neq 0$, implying that $F(f) \neq 0$ for f in a neighborhood to 0.) ■

REMARK 5.8 Classical counter-examples in the context of the *direct method* of calculus of variations show that the step of verifying the existence of a minimizer should not be taken too lightly. For instance, the functional

$$J(u) := \int_0^1 \left[(u'(s)^2 - 1)^2 + u(s)^2\right] \mathrm{d}s$$

does not have a minimizer in H_0^1, but J can be made arbitrarily close to 0 by choosing piecewise-linear functions u with slope $|u'| = 1$ oscillating around 0. We refer to any text book on calculus of variations. In the situation above, local 'convexity' in the sense of a positive definite second derivative prevents this phenomenon. An alternative method of proof for the existence of a minimizer is to show that J is (lower semi-) continuous in the weak sense.

5.1.2. The general case. In the general case (cf. Proposition 5.2), we define the function $H : H_0^1 \times \mathbb{R} \to H_0^1$ by

$$\begin{aligned} &H(f,x)(t) \\ &:= f_t - \frac{\rho\left(x - \rho G(f)\right)}{\bar\rho^2 F(f)} \frac{\left\{\left\langle \sigma\left(K\dot f\right), \mathbf{1}_{[0,t]}\right\rangle + \left\langle \sigma'\left(K\dot f\right)\dot f, K\mathbf{1}_{[0,t]}\right\rangle\right\}}{\bar\rho^2 F(f)} \\ &\quad + \frac{\left(x - \rho G(f)\right)^2}{\bar\rho^2 F^2(f)} \left\langle \left(\sigma\sigma'\right)\left(K\dot f\right), K\mathbf{1}_{[0,t]}\right\rangle \\ &= f_t - \frac{\rho\left(x - \rho G(f)\right)}{\bar\rho^2 F(f)} \left(R_2(f)(t) + R_3(f)(t)\right) \\ &\quad + \frac{\left(x - \rho G(f)\right)^2}{\bar\rho^2 F(f)^2} R_1(f)(t), \end{aligned} \tag{28}$$

where $R_2, R_3 : H_0^1 \to H_0^1$ are defined by

$$R_2(f)(t) := \left\langle \sigma(K\dot f), \mathbf{1}_{[0,t]}\right\rangle, \tag{29}$$

$$R_3(f)(t) := \left\langle \sigma'(K\dot f)\dot f, K\mathbf{1}_{[0,t]}\right\rangle, \tag{30}$$

$t \in [0,1]$.

One easily checks that G, R_2, R_3 are smooth in the Fréchet sense.

LEMMA 5.9 *The functions* $G : H_0^1 \to \mathbb{R}$, $R_2 : H_0^1 \to H_0^1$ *and* $R_3 : H_0^1 \to H_0^1$ *are smooth in Fréchet sense.*

Proof The proof of smoothness is clear. We report the actual derivatives. For G we get

$$D^N G(f) \cdot (g_1, \ldots, g_N) = \left\langle \sigma^{(N)}\left(K\dot f\right)\dot f, \prod_{i=1}^N K\dot g_i \right\rangle + \sum_{k=1}^N \left\langle \sigma^{(N-1)}\left(K\dot f\right), \dot g_k \prod_{i\neq k} K\dot g_i \right\rangle.$$

For R_2 and, respectively, R_3, we obtain

$$\left(D^N R_2(f) \cdot (g_1, \ldots, g_N)\right)(t) = \int_0^t \sigma^{(N)}\left((K\dot f)(s)\right) \prod_{i=1}^N (K\dot g_i)(s)\, \mathrm{d}s,$$

and

$$\begin{aligned}&\left(D^N R_3(f)\cdot(g_1,\dots,g_N)\right)(t)\\&=\left\langle \sigma^{(N+1)}\left(K\dot f\right)\dot f K\mathbf{1}_{[0,t]},\prod_{i=1}^N K\dot g_i\right\rangle\\&\quad+\sum_{k=1}^N\left\langle \sigma^{(N)}\left(K\dot f\right)K\mathbf{1}_{[0,t]},\dot g_k\prod_{i\neq k}K\dot g_i\right\rangle.\end{aligned}$$

■

THEOREM 5.10 *Let σ be smooth with $\sigma(0)\neq 0$. Then the energy $I(x)$ as defined in* (24) *is smooth in a neighborhood of $x=0$.*

Proof The proof is similar to the proof of Theorem 5.7. In fact, the only difference is in establishing invertibility of $DH(0,0)$ and the existence of a minimizer.

Note that (28) contains three terms. The derivative of the first term ($f\mapsto f$) is always equal to $\mathrm{id}_{H_0^1}$. For the second term, we note that

$$(x-\rho G(f))|_{x=0,\,f=0}=0.$$

Hence, the only non-vanishing contribution to the derivative of the second term evaluated in direction $g\in H_0^1$ at $x=0$, $f=0$ and $t\in[0,1]$ is

$$\begin{aligned}&\frac{\rho^2 DG(0)\cdot g}{\bar\rho^2 F(0)}(R_2(0)+R_3(0))=\frac{\rho^2\sigma_0 g(1)}{\bar\rho^2\sigma_0^2}(\sigma_0 t+0)\\&=\frac{\rho^2}{\bar\rho^2}g(1)t.\end{aligned}$$

For the same reason, the derivative of the third term at $(f,x)=(0,0)$ vanishes entirely. Hence,

$$(DH(0,0)\cdot g)(t)=g(t)+\frac{\rho^2}{\bar\rho^2}g(1)t.$$

It is easy to see that $g\mapsto DH(0,0)\cdot g$ is invertible. Indeed, let us construct the pre-image $g=DH(0,0)^{-1}\cdot h$ of some $h\in H_0^1$. At $t=1$ we have

$$\frac{\bar\rho^2+\rho^2}{\bar\rho^2}g(1)=h(1),$$

implying $g(1)=\bar\rho^2 h(1)$. For $0\le t<1$, we then get

$$\begin{aligned}&g(t)+\frac{\rho^2}{\bar\rho^2}g(1)t=g(t)+\frac{\rho^2}{\bar\rho^2}\bar\rho^2 h(1)t\\&=g(t)+\rho^2 h(1)t=h(t),\end{aligned}$$

or $g(t)=h(t)-\rho^2 h(1)t$.

For existence of the minimizer, note that

$$D^2J_0(0)\cdot(g,g)=\frac{\rho^2}{\bar\rho^2}g(1)^2+\|g\|^2_{H_0^1},$$

which is again positive definite. ■

REMARK 5.11 Though only formulated in terms of 'smoothness', it is easy to show that $\sigma\in C^k$ implies that $I\in C^{k-1}$ (locally at 0).

5.2. *Energy expansion*

Having established smoothness of the energy I as well as of the minimizing configuration $x\mapsto f^x$ locally around $x=0$, we can proceed with computing the Taylor expansion of f^x around $x=0$. We will once more rely on the first order optimality condition given in Proposition 5.2. Plugging the Taylor expansion of f^x into $\mathcal{I}_x$ will then give us the local Taylor expansion of $I(x)$.

5.2.1. Expansion of the minimizing configuration.

THEOREM 5.12 *We have*

$$\begin{aligned}f_t^x&=\alpha_t x+\beta_t\frac{x^2}{2}+\mathcal{O}\left(x^3\right),\\\alpha_t&=\frac{\rho}{\sigma_0}t,\\\beta_t&=2\frac{\sigma_0'}{\sigma_0^3}\left[\rho^2\left\langle K1\,,\mathbf{1}_{[0,t]}\right\rangle+\left\langle K\mathbf{1}_{[0,t]}\,,1\right\rangle-3\rho^2 t\left\langle K1\,,1\right\rangle\right].\end{aligned}$$

REMARK 5.13 Non-Markovian transversality In the RL-fBM case, $K(t,s)=\sqrt{2H}|t-s|^\gamma$ with $\gamma=H-1/2$ one computes

$$\left\langle 1,K1_{[0,t]}\right\rangle=\frac{1}{(1+\gamma)(2+\gamma)}\left\{1-(1-t)^{2+\gamma}\right\}\in C^1[0,1].$$

Interestingly, the transversality condition known from the Markovian setting ($q_1=0$, which readily translates to $\dot f_1^x=0$ there) remains valid here (for $\rho=0$), at least to order x^2, in the sense that

$$\dot f_t^x\approx\beta_t\frac{x^2}{2}=(\text{const})\,(1-t)^{1+\gamma}\,|_{t=1}=0$$

Proof of Theorem 5.12 *First order expansion*:

Up to the order needed in order to get the first order term, we have

$$\begin{aligned}f_t^x&=\alpha_t x+\mathcal{O}(x^2),\\\dot f_t^x&=\dot\alpha_t x+\mathcal{O}(x^2),\\\sigma(K\dot f^x)&=\sigma_0+\sigma_0' K\dot\alpha x+\mathcal{O}(x^2),\\\sigma'(K\dot f^x)&=\sigma_0'+\sigma_0'' K\dot\alpha x+\mathcal{O}(x^2),\\F(f^x)&=\langle\sigma^2(K\dot f^x),1\rangle\\&=\sigma_0^2+\mathcal{O}(x),\\G(f^x)&=\langle\sigma(K\dot f^x),\dot f^x\rangle\\&=\langle\sigma_0\,,\dot\alpha\rangle\,x+\mathcal{O}(x^2).\end{aligned}$$

Therefore,

$$\begin{aligned}
\langle \sigma(K\dot{f}^x), \mathbf{1}_{[0,t]} \rangle &= \sigma_0 t + \mathcal{O}(x),\\
\langle \sigma'(K\dot{f}^x)\dot{f}^x, K\mathbf{1}_{[0,t]} \rangle &= \mathcal{O}(x),\\
\langle \sigma\sigma'(K\dot{f}^x), K\mathbf{1}_{[0,t]} \rangle &= \mathcal{O}(1),\\
x - \rho G(f^x) &= (1-\rho\sigma_0\alpha_1)x + \mathcal{O}(x^2),\\
(x-\rho G(f^x))^2 &= \mathcal{O}(x^2).
\end{aligned}$$

This yields for the first order term in (25)

$$\alpha_t = \frac{\rho(1-\rho\sigma_0\alpha_1)}{\bar{\rho}^2\sigma_0} t.$$

Setting $t=1$, we get

$$\alpha_1 = \frac{\rho}{\bar{\rho}^2\sigma_0} - \frac{\rho^2}{\bar{\rho}^2}\alpha_1,$$

which is solved by $\alpha_1 = \rho/\sigma_0$. Inserting this term back into the equation for α_t, we get

$$\alpha_t = \frac{\rho}{\sigma_0} t. \tag{31}$$

Second order expansion:

Using (31) and the ansatz $f_t^x = \alpha_t x + \frac{1}{2}\beta_t x^2 + \mathcal{O}(x^3)$, we re-compute the relevant terms appearing in the (25). We have

$$\sigma(K\dot{f}^x(s)) = \sigma_0 + \sigma_0' \frac{\rho}{\sigma_0}(K1)(s)x + \mathcal{O}(x^2)$$

and analogously for σ replaced by σ', $\sigma\sigma'$. This implies

$$\begin{aligned}
\left\langle \sigma(K\dot{f}^x), \mathbf{1}_{[0,t]} \right\rangle &= \sigma_0 t + \sigma_0' \frac{\rho}{\sigma_0} \left\langle K1, \mathbf{1}_{[0,t]} \right\rangle x + \mathcal{O}(x^2),\\
\left\langle \sigma'(K\dot{f}^x)\dot{f}^x, K\mathbf{1}_{[0,t]} \right\rangle &= \rho \frac{\sigma'}{\sigma_0} \left\langle K\mathbf{1}_{[0,t]}, 1 \right\rangle x + \mathcal{O}(x^2),\\
\left\langle \sigma\sigma'(K\dot{f}^x), K\mathbf{1}_{[0,t]} \right\rangle &= \sigma_0\sigma_0' \left\langle K\mathbf{1}_{[0,t]}, 1 \right\rangle + \mathcal{O}(x).
\end{aligned}$$

Using the notation introduced earlier, we have

$$\begin{aligned}
F(f^x) &= \sigma_0^2 + 2\sigma_0'\rho \langle K1, 1\rangle x + \mathcal{O}(x^2),\\
G(f^x) &= \rho x + \left(\frac{1}{2}\sigma_0\beta_1 + \rho^2 \frac{\sigma_0'}{\sigma_0^2} \langle K1, 1\rangle\right) x^2 + \mathcal{O}(x^3).
\end{aligned}$$

This directly implies

$$\begin{aligned}
x - \rho G(f^x) &= \bar{\rho}^2 x - \rho\left(\frac{1}{2}\sigma_0\beta_1 + \rho^2\frac{\sigma_0'}{\sigma_0^2}\langle K1,1\rangle\right)x^2 + \mathcal{O}(x^3),\\
(x-\rho G(f^x))^2 &= \bar{\rho}^4 x^2 - 2\bar{\rho}^2\rho\\
&\times \left(\frac{1}{2}\sigma_0\beta_1 + \rho^2\frac{\sigma_0'}{\sigma_0^2}\langle K1,1\rangle\right)x^3 + \mathcal{O}(x^4).
\end{aligned}$$

We next compute some auxiliary terms appearing in (25).

$$\begin{aligned}
N_1 &\coloneqq \rho(x-\rho G(f^x))\left(\left\langle \sigma(K\dot{f}^x), \mathbf{1}_{[0,t]}\right\rangle + \left\langle \sigma'(K\dot{f}^x)\dot{f}^x, K\mathbf{1}_{[0,t]}\right\rangle\right)\\
&= \rho\bar{\rho}^2\sigma_0 t x + \left[\rho^2\bar{\rho}^2\frac{\sigma_0'}{\sigma_0}\left(\left\langle K1, \mathbf{1}_{[0,t]}\right\rangle + \left\langle K\mathbf{1}_{[0,t]}, 1\right\rangle\right)\right.\\
&\quad \left. -\rho^4\frac{\sigma_0'}{\sigma_0} t\langle K1,1\rangle - \frac{1}{2}\rho^2\sigma_0^2 t\beta_1\right]x^2 + \mathcal{O}(x^3)
\end{aligned}$$

The corresponding denominator is $\bar{\rho}^2 F(f^x)$. Using the formula

$$\frac{a_1 x + a_2 x^2 + \mathcal{O}(x^3)}{b_0 + b_1 x + \mathcal{O}(x^2)} = \frac{a_1}{b_0}x + \frac{a_2 b_0 - a_1 b_1}{b_0^2}x^2 + \mathcal{O}(x^3),$$

we obtain

$$\begin{aligned}
\frac{N_1}{\bar{\rho}^2 F(f^x)} &= \frac{\rho}{\sigma_0} t x + \left[\rho^2\frac{\sigma_0'}{\sigma_0^3}\left(\left\langle K1, \mathbf{1}_{[0,t]}\right\rangle + \left\langle K\mathbf{1}_{[0,t]}, 1\right\rangle\right)\right.\\
&\quad \left. -\left(\frac{\rho^4}{\bar{\rho}^2} + 2\rho^2\right)\frac{\sigma_0'}{\sigma_0^3} t\langle K1,1\rangle - \frac{1}{2}\frac{\rho^2}{\bar{\rho}^2}\beta_1 t\right]x^2 + \mathcal{O}(x^3)
\end{aligned} \tag{32}$$

For the second term in (25), let

$$\begin{aligned}
N_2 &\coloneqq (x-\rho G(f^x))^2 \left\langle (\sigma\sigma')(K\dot{f}^x), K\mathbf{1}_{[0,t]}\right\rangle\\
&= \bar{\rho}^4\sigma_0\sigma_0' \left\langle K\mathbf{1}_{[0,t]}, 1\right\rangle x^2 + \mathcal{O}(x^3).
\end{aligned}$$

The corresponding denominator is $\bar{\rho}^2 F(f^x)^2 = \bar{\rho}^2\sigma_0^4 + \mathcal{O}(x)$. Hence,

$$\frac{N_2}{\bar{\rho}^2 F(f^x)^2} = \bar{\rho}^2\frac{\sigma_0'}{\sigma_0^3}\left\langle K\mathbf{1}_{[0,t]}, 1\right\rangle x^2 + \mathcal{O}(x^3). \tag{33}$$

Combining (32) and (33), we get

$$\begin{aligned}
f_t^x &= \frac{\rho}{\sigma_0} t x + \left[\rho^2\frac{\sigma_0'}{\sigma_0^3}\left(\left\langle K1, \mathbf{1}_{[0,t]}\right\rangle + \left\langle K\mathbf{1}_{[0,t]}, 1\right\rangle\right)\right.\\
&\quad -\frac{\rho^4}{\bar{\rho}^2}\frac{\sigma_0'}{\sigma_0^3} t\langle K1,1\rangle\\
&\quad \left. -\frac{1}{2}\frac{\rho^2}{\bar{\rho}^2}\beta_1 t - 2\rho^2\frac{\sigma_0'}{\sigma_0^3} t\langle K1,1\rangle + \bar{\rho}^2\frac{\sigma_0'}{\sigma_0^3}\left\langle K\mathbf{1}_{[0,t]}, 1\right\rangle\right]x^2\\
&\quad + \mathcal{O}(x^3)
\end{aligned}$$

We shall next compute β_1. Taking the second order terms on both sides and letting $t=1$, we obtain

$$\begin{aligned}
\frac{1}{2}\beta_1 &= \rho^2\frac{\sigma_0'}{\sigma_0^3} 2\langle K1,1\rangle - \frac{\rho^4}{\bar{\rho}^2}\frac{\sigma_0'}{\sigma_0^3}\langle K1,1\rangle\\
&\quad - \frac{1}{2}\frac{\rho^2}{\bar{\rho}^2}\beta_1 - 2\rho^2\frac{\sigma_0'}{\sigma_0^3}\langle K1,1\rangle + \bar{\rho}^2\frac{\sigma_0'}{\sigma_0^3}\langle K1,1\rangle.
\end{aligned}$$

Moving β_1 to the other side with $1 + \rho^2/\bar{\rho}^2 = 1/\bar{\rho}^2$ and collecting terms on the right hand side, we arrive at

$$\begin{aligned}
\frac{1}{2}\frac{1}{\bar{\rho}^2}\beta_1 &= \frac{\sigma_0'}{\sigma_0^3}\langle K1,1\rangle\left(2\rho^2 - \frac{\rho^4}{\bar{\rho}^2} - 2\rho^2 + \bar{\rho}^2\right)\\
&= \frac{1-2\rho^2}{\bar{\rho}^2}\frac{\sigma_0'}{\sigma_0^3}\langle K1,1\rangle
\end{aligned}$$

We conclude that

$$\beta_1 = 2(1-2\rho^2)\frac{\sigma_0'}{\sigma_0^3}\langle K1,1\rangle$$

Hence, we obtain

$$\beta_t = 2\frac{\sigma_0'}{\sigma_0^3}\left[\rho^2\left\langle K1\,,\mathbf{1}_{[0,t]}\right\rangle + \left\langle K\mathbf{1}_{[0,t]}\,,1\right\rangle - 3\rho^2 t\left\langle K1\,,1\right\rangle\right].$$

■

5.2.2. Energy expansion in the general case. Now we compute the Taylor expansion of $I(x)$ as defined in Proposition 5.1. We start with the second term. Plugging in the optimal path $f_t^x = \alpha_t x + \frac{1}{2}\beta_t x^2 + \mathcal{O}(x^3)$ (and using $\langle\dot\beta\,,1\rangle = \beta_1$ as $\beta_0 = 0$) we obtain

$$\frac{1}{2}\left\langle \dot f^x\,,\dot f^x\right\rangle = \frac{1}{2}\frac{\rho^2}{\sigma_0^2}x^2 + \frac{1}{2}\frac{\rho}{\sigma_0}\beta_1 x^3 + \mathcal{O}(x^4).$$

Inserting $\beta_1 = 2(1-2\rho^2)(\sigma_0'/\sigma_0^3)\left\langle K1\,,1\right\rangle$ into the above formula for $(x-\rho G(f^x))^2$, we get

$$(x-\rho G(f^x))^2 = \bar\rho^4 x^2 - 2\bar\rho^4\rho\frac{\sigma_0'}{\sigma_0^2}\left\langle K1\,,1\right\rangle x^3 + \mathcal{O}(x^4).$$

Recall the denominator

$$2\bar\rho^2 F(f^x) = 2\bar\rho^2\sigma_0^2 + 4\bar\rho^2\sigma_0'\rho\left\langle K1\,,1\right\rangle x + \mathcal{O}(x^2).$$

Using the expansion of a fraction

$$\frac{a_2x^2 + a_3x^3 + \mathcal{O}(x^4)}{b_0 + b_1x + \mathcal{O}(x^2)} = \frac{a_2}{b_0}x^2 + \frac{a_3b_0 - a_2b_1}{b_0^2}x^3 + \mathcal{O}(x^4),$$

we obtain from

$$\begin{aligned}\frac{(x-\rho G(f^x))^2}{2\bar\rho^2F(f^x)} &= \frac{\bar\rho^4}{2\bar\rho^2\sigma_0^2}x^2\\ &\quad + \frac{\left(-2\bar\rho^4\rho\frac{\sigma_0'}{\sigma_0^2}\left\langle K1\,,1\right\rangle\right)2\bar\rho^2\sigma_0^2 - \bar\rho^4\left(4\bar\rho^2\sigma_0'\rho\left\langle K1\,,1\right\rangle\right)}{4\bar\rho^4\sigma_0^4}x^3\\ &\quad + \mathcal{O}(x^4)\\ &= \frac{\bar\rho^2}{2\sigma_0^2}x^2 - 2\bar\rho^2\rho\frac{\sigma_0'}{\sigma_0^4}\left\langle K1\,,1\right\rangle x^3 + \mathcal{O}(x^4).\end{aligned}$$

We note that

$$\begin{aligned}&\frac{1}{2}\frac{\rho}{\sigma_0}\beta_1 - 2\bar\rho^2\rho\frac{\sigma_0'}{\sigma_0^4}\left\langle K1\,,1\right\rangle\\ &= \left((1-2\rho^2) - 2(1-\rho^2)\right)\rho\frac{\sigma_0'}{\sigma_0^4}\left\langle K1\,,1\right\rangle = -\rho\frac{\sigma_0'}{\sigma_0^4}\left\langle K1\,,1\right\rangle.\end{aligned}$$

Adding both terms, we arrive at the

PROPOSITION 5.14 *The energy expansion to third order gives*

$$I(x) = \frac{1}{2\sigma_0^2}x^2 - \rho\frac{\sigma_0'}{\sigma_0^4}\left\langle K1\,,1\right\rangle x^3 + \mathcal{O}(x^4).$$

5.2.3. Energy expansion for the Riemann-Liouville kernel. Let us specialize the energy expansion given in Proposition 5.14 for the Riemann-Liouville fBm. Choose $\gamma = H - \frac{1}{2}$ and recall that the kernel K takes the form $K(t,s) = (t-s)^\gamma$. We get

$$(K1)(t) = \int_0^t K(t,s)\,\mathrm{d}s = \int_0^t (t-s)^\gamma\,\mathrm{d}s = \frac{t^{1+\gamma}}{1+\gamma}.$$

The key term $\langle K1\,,1\rangle$ appearing in the energy expansion now gives

$$\begin{aligned}\left\langle K1\,,1\right\rangle &= \int_0^1 (K1)(t)\,\mathrm{d}t = \int_0^1 \frac{t^{1+\gamma}}{1+\gamma}\mathrm{d}t = \frac{1}{(1+\gamma)(2+\gamma)}\\ &= \frac{1}{(H+1/2)(H+3/2)}.\end{aligned}$$

Plugging the above formula into the energy expansion, we obtain the energy expansion for the Riemann-Liouville fractional Browian motion

$$I(x) = \frac{1}{2\sigma_0^2}x^2 - \frac{\rho}{(H+1/2)(H+3/2)}\frac{\sigma_0'}{\sigma_0^4}x^3 + \mathcal{O}(x^4).$$

For completeness, let us also fully describe the time-dependence of the second order term β_t in the expansion of the optimal trajectory f_t^x. Unlike the first order time, here we do not have a linear movement any more. Indeed

$$\begin{aligned}\left\langle K1\,,\mathbf{1}_{[0,t]}\right\rangle &= \int_0^t (K1)(s)\,\mathrm{d}s = \int_0^t \frac{s^{1+\gamma}}{1+\gamma}\mathrm{d}s\\ &= \frac{t^{2+\gamma}}{(1+\gamma)(2+\gamma)},\end{aligned} \tag{34}$$

$$\left\langle K\mathbf{1}_{[0,t]}\,,1\right\rangle = \frac{1}{(1+\gamma)(2+\gamma)}\left(1-(1-t)^{2+\gamma}\right). \tag{35}$$

6. Proof of the pricing formula

Fix $x \geq 0$ and $\hat x = (\varepsilon/\hat\varepsilon)x$ where $\varepsilon = t^{1/2}$ and $\hat\varepsilon = t^H = \varepsilon^{2H}$. We have

$$\begin{aligned}c(\hat x, t) &= E\left(\exp(X_t) - \exp\hat x\right)^+\\ &= E\left(\exp\left(X_1^\varepsilon\right) - \exp\hat x\right)^+\\ &= E\left(\exp\left(\frac{\varepsilon}{\hat\varepsilon}\hat X_1^\varepsilon\right) - \exp\left(\frac{\varepsilon}{\hat\varepsilon}x\right)\right)^+\end{aligned}$$

where we recall

$$\begin{aligned}\hat X_1^\varepsilon \equiv \frac{\hat\varepsilon}{\varepsilon}X_1^\varepsilon &= \int_0^1 \sigma(\hat\varepsilon\hat B)\hat\varepsilon\,\mathrm{d}\left(\bar\rho W + \rho B\right)\\ &\quad - \frac{1}{2}\varepsilon\hat\varepsilon\int_0^1 \sigma\left(\hat\varepsilon\hat B_t\right)^2\mathrm{d}t.\end{aligned}$$

Consider a Cameron-Martin perturbation of $\hat{X}_1^\varepsilon$. That is, for a Cameron-Martin path $\mathrm{h} = (h,f) \in H_0^1 \times H_0^1$ consider a measure change corresponding to a transformation $\hat{\varepsilon}(W,B) \rightsquigarrow \hat{\varepsilon}(W,B) + (h,f)$ (transforming the Brownian motions to Brownian motions with drift), we obtain the Girsanov density

$$G_\varepsilon = \exp\left(-\frac{1}{\hat{\varepsilon}}\int_0^1 \dot{h}_s \,\mathrm{d}W_s - \frac{1}{\hat{\varepsilon}}\int_0^1 \dot{f}_s \,\mathrm{d}B_s - \frac{1}{2\hat{\varepsilon}^2}\int_0^1 \left(\dot{h}_s^2 + \dot{f}_s^2\right)\mathrm{d}s\right). \tag{36}$$

Under the new measure, $\hat{X}_1^\varepsilon$ becomes $\hat{Z}_1^\varepsilon$, where

$$\hat{Z}_1^\varepsilon = \int_0^1 \sigma(\hat{\varepsilon}\hat{B}_t + \hat{f}_t)\left[\hat{\varepsilon}\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) + \mathrm{d}(\bar{\rho}h_t + \rho f_t)\right] - \frac{1}{2}\varepsilon\hat{\varepsilon}\int_0^1 \sigma(\hat{\varepsilon}\hat{B}_t + \hat{f}_t)^2\,\mathrm{d}t.$$

Definition 6.1 For fixed $x \geq 0$, write $(h,f) \in \mathcal{K}^x$ if $\Phi_1(h,f,\hat{f}) = x$. Call such (h,f) admissible for arrival at log-strike x. Call (h^x,f^x) the cheapest admissible control, which attains

$$I(x) = \inf_{h,f\in H_0^1}\left\{\frac{1}{2}\int_0^1 \dot{h}^2\,\mathrm{d}t + \frac{1}{2}\int_0^1 \dot{f}^2\,\mathrm{d}t : \Phi_1\left(h,f,\hat{f}\right) = x\right\},$$

where we recall that $\hat{f} = K\dot{f}$ and

$$\Phi_1(h,f,\hat{f}) = \int_0^1 \sigma(\hat{f})\,\mathrm{d}(\bar{\rho}h + \rho f).$$

For any Cameron-Martin path (h,f), the perturbed random variable $\hat{Z}_1^\varepsilon$ admits a stochastic Taylor expansion with respect to $\hat{\varepsilon}$.

Lemma 6.2 *Fix $(h,f) \in \mathcal{K}^x$ and define $\hat{Z}_1^\varepsilon$ accordingly. Then*

$$\hat{Z}_1^\varepsilon = x + \hat{\varepsilon}g_1 + \hat{\varepsilon}^2 R_2^\varepsilon, \tag{37}$$

where g_1 is a Gaussian random variable, given explicitly by

$$g_1 = \int_0^1 \{\sigma(\hat{f}_t)\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) + \sigma'(\hat{f}_t)\hat{B}_t\,\mathrm{d}(\bar{\rho}h_t + \rho f_t)\}, \tag{38}$$

and

$$R_2^\varepsilon = \int_0^1 \sigma'\left(\hat{f}_t\right)\hat{B}_t\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) - \frac{1}{2}\frac{\varepsilon}{\hat{\varepsilon}}\int_0^1 \sigma(\hat{\varepsilon}\hat{B}_t + \hat{f}_t)^2\,\mathrm{d}t + \frac{1}{2\hat{\varepsilon}^2}\int_0^{\hat{\varepsilon}}\int_0^1 \sigma''\left(\zeta\hat{B}_t + \hat{f}_t\right)\hat{B}_t^2 \times\left[\hat{\varepsilon}\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) + \mathrm{d}(\bar{\rho}h_t + \rho f_t)\right](\hat{\varepsilon} - \zeta)\,\mathrm{d}\zeta. \tag{39}$$

Proof By a stochastic Taylor expansion for the controlled process $\hat{Z}_t^\varepsilon$ with control $(h,f) \in \mathcal{K}^x$ as in Definition 6.1 and thanks to $\sigma \in C^2$, we have at $t = 1$

$$\begin{aligned}\hat{Z}_1^\varepsilon &= \int_0^1 \sigma(\hat{\varepsilon}\hat{B} + \hat{f})\left[\hat{\varepsilon}\,\mathrm{d}(\bar{\rho}W + \rho B) + \mathrm{d}(\bar{\rho}h + \rho f)\right] - \frac{1}{2}\varepsilon\hat{\varepsilon}\int_0^1 \sigma(\hat{\varepsilon}\hat{B}_t + \hat{f}_t)^2\,\mathrm{d}t\\ &= \int_0^1 \sigma(\hat{f})\,\mathrm{d}(\bar{\rho}h + \rho f) + \hat{\varepsilon}\int_0^1 \{\sigma(\hat{f})\,\mathrm{d}(\bar{\rho}W + \rho B) + \sigma'(\hat{f})\hat{B}\,\mathrm{d}(\bar{\rho}h + \rho f)\}\\ &\quad + \hat{\varepsilon}^2\int_0^1 \sigma'\left(\hat{f}_t\right)\hat{B}_t\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) - \frac{1}{2}\varepsilon\hat{\varepsilon}\int_0^1 \sigma(\hat{\varepsilon}\hat{B}_t + \hat{f}_t)^2\,\mathrm{d}t\\ &\quad + \frac{1}{2}\int_0^{\hat{\varepsilon}}\int_0^1 \sigma''\left(\zeta\hat{B}_t + \hat{f}_t\right)\hat{B}_t^2 \times\left[\hat{\varepsilon}\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) + \mathrm{d}(\bar{\rho}h_t + \rho f_t)\right](\hat{\varepsilon} - \zeta)\,\mathrm{d}\zeta.\end{aligned}$$

Collecting terms in powers of $\hat{\varepsilon}$ and with the random variable g_1 as in (38) (recalling that $\hat{\varepsilon}\varepsilon \in \mathcal{O}(\hat{\varepsilon}^2)$), we have

$$\hat{Z}_1^\varepsilon = \int_0^1 \sigma(\hat{f})\,\mathrm{d}(\bar{\rho}h + \rho f) + \hat{\varepsilon}g_1 + \mathcal{O}(\hat{\varepsilon}^2),$$

furthermore, since $(h,f) \in \mathcal{K}^x$, by the definition of Φ_1, it holds that

$$\int_0^1 \sigma(\hat{f})\,\mathrm{d}(\bar{\rho}h + \rho f) = x.$$

This proves the statement (37) and the statement that g_1 is Gaussian is immediate from the form (38). ■

Finally, we determine an explicit form of the Girsanov density G_ε for the choice where (h^x,f^x) in (36) are chosen the cheapest admissible control (cf. Definition 6.1. Similarly to classical works of Azencott, Ben Arous and others, see, for instance, Ben Arous (1988), we show that the stochastic integrals in the exponent of G_ε are proportional to the first order term g_1 (with factor $I'(x)$) *when evaluated at the minimizing configuration* (h^x,f^x).

Lemma 6.3 *We have*

$$\int_0^1 \dot{h}_t^x\,\mathrm{d}W_t + \int_0^1 \dot{f}_t^x\,\mathrm{d}B_t = I'(x)\,g_1.$$

Proof See Lemma A.2. ■

With these preparations in place, we are now ready to prove the pricing formula from Section 3.

Proof of Theorem 3.2 With a Girsanov factor (all integrals on $[0,1]$)

$$G_\varepsilon = \mathrm{e}^{-1/\hat{\varepsilon}\int \dot{h}\,\mathrm{d}W - \frac{1}{\varepsilon}\int \dot{f}\,\mathrm{d}B - \frac{1}{2\hat{\varepsilon}^2}\int(\dot{h}^2 + \dot{f}^2)\mathrm{d}t}$$

and (evaluated at the minimizer)

$$G_\varepsilon|_* = \mathrm{e}^{-I(x)/\hat{\varepsilon}^2}\,\mathrm{e}^{-I'(x)g_1(\omega)/\hat{\varepsilon}},$$

we have, setting $\hat{U}^\varepsilon := \hat{Z}_1^\varepsilon - x = \hat{\varepsilon} g_1 + \hat{\varepsilon}^2 R_2^\varepsilon$

$$\begin{aligned} c(\hat{x},t) &= E\left[\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{Z}_1^\varepsilon\right) - \exp\left(\frac{\varepsilon}{\hat{\varepsilon}}x\right)\right)^+ G_\varepsilon|_*\right] \\ &= \mathrm{e}^{\varepsilon/\hat{\varepsilon}x} E\left[\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{U}^\varepsilon\right) - 1\right)^+ G_\varepsilon|_*\right] \\ &= \mathrm{e}^{-I(x)/\hat{\varepsilon}^2}\,\mathrm{e}^{\varepsilon/\hat{\varepsilon}x} E\left[\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{U}^\varepsilon\right) - 1\right)^+ \mathrm{e}^{-I'(x)g_1/\hat{\varepsilon}}\right] \\ &= \mathrm{e}^{-I(x)/\hat{\varepsilon}^2}\,\mathrm{e}^{\varepsilon/\hat{\varepsilon}x} E\left[\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{U}^\varepsilon\right) - 1\right)\mathrm{e}^{-(I'(x)/\hat{\varepsilon}^2)\hat{U}^\varepsilon}\right. \\ &\quad \left.\times\, \mathrm{e}^{I'(x)R_2^\varepsilon}\mathbf{1}_{\hat{U}^\varepsilon \geq 0}\right]. \\ &= \mathrm{e}^{-I(x)/\hat{\varepsilon}^2}\,\mathrm{e}^{\varepsilon/\hat{\varepsilon}x} J(\varepsilon,x)\,. \end{aligned}$$

■

7. Proof of the moderate deviation expansions

In Section 2, we pointed out that (iiic) is exactly what one gets from (call price) large deviations (8), if heuristically applied to $x\varepsilon^{2\beta}$. We now give a proper derivation based on moderate deviations.

LEMMA 7.1 *Assume* (iiia-b) *from Assumption* 2.4. *Then an upper moderate deviation estimate holds both for calls and digital calls. That is, we have*

(iiic) *For every* $\beta \in (0,H)$, *and every fixed* $x > 0$, *and* $\hat{x}_\varepsilon := x\varepsilon^{1-2H+2\beta}$,

$$E[(\mathrm{e}^{X_1^\varepsilon} - \mathrm{e}^{\hat{x}_\varepsilon})^+] \leq \exp\left(-\frac{x^2 + o(1)}{2\sigma_0^2\varepsilon^{4H-4\beta}}\right)$$

and also

$$P[X_1^\varepsilon > \hat{x}_\varepsilon] \leq \exp\left(-\frac{x^2 + o(1)}{2\sigma_0^2\varepsilon^{4H-4\beta}}\right). \tag{40}$$

Proof Recall $\sigma(.)$ smooth but unbounded and recall $\hat{x}_\varepsilon := x\varepsilon^{1-2H+2\beta}$. In case of $\beta = 0$ and $H = 1/2$ a large deviation principle (LDP) for $(X_1^\varepsilon\hat{\varepsilon}/\varepsilon)$ is readily reduced, via exponential equivalence, to a LDP for the family of stochastic Itô integrals given by $\int \sigma(\hat{\varepsilon}\hat{B})\hat{\varepsilon}\,\mathrm{d}Z$ for some Brownian Z, ρ-correlated with B. There are then many ways to establish a LDP for this family. A particularly convenient one, that requires no growth restriction on σ, uses continuity of stochastic integration with respect to the rough path $(B,Z,\int B\,\mathrm{d}Z) = (B,Z,\int \hat{B}\,\mathrm{d}Z)$ in suitable metrics, for which a LDP is known (Friz and Hairer 2014, Ch 9.3). It was pointed out in Bayer *et al.* (2017) that a similar reasoning is possible when $H < 1/2$, the rough path is then replaced by a 'richer enhancement' of (B,Z), the precise size of which depends on H, for which again one has a LDP. A moderate deviation priniple (MDP) for $(X_1^\varepsilon\hat{\varepsilon}/\varepsilon)$ is a LDP for $(\varepsilon^{-2\beta}X_1^\varepsilon\hat{\varepsilon}/\varepsilon)$ for $\beta \in (0,H)$. This can be reduced to a LDP, with $\bar{\varepsilon} := \varepsilon^{-2\beta}\hat{\varepsilon} = \varepsilon^{2H-2\beta}$, for

$$\varepsilon^{-2\beta}\int_0^1 \sigma(\hat{\varepsilon}\hat{B})\hat{\varepsilon}\,\mathrm{d}Z = \int_0^1 \sigma(\hat{\varepsilon}\hat{B})\bar{\varepsilon}\,\mathrm{d}Z \equiv \int_0^1 \sigma_\varepsilon(\bar{\varepsilon}\hat{B})\bar{\varepsilon}\,\mathrm{d}Z$$

with speed $\bar{\varepsilon}^2$. Since $\sigma_\varepsilon(\cdot) \equiv \sigma(\varepsilon^{2\beta}\cdot)$ converges (with all derivatives) locally uniformly to the constant function σ_0, and one checks that the above is exponentially equivalent to the (Gaussian) family given by $\sigma_0\bar{\varepsilon}Z_1$, with law $\mathcal{N}(0,\sigma_0^2\bar{\varepsilon}^2) = \mathcal{N}(0,\sigma_0^2\varepsilon^{4H-4\beta})$ which gives (40), even with equality. (By localization, exponential equivalence can again be done for σ without growth restrictions.)

We have not yet used either assumption (iiia-b). These become important in order to extend estimate (40) to the case of genuine call payoffs. We can follow here a well-known argument (e.g. Forde and Jacquier 2009; Pham 2010; Forde and Zhang 2017) with the 'moderate' caveat to carry along a factor $\varepsilon^{2\beta}$. In fact, this is follows precisely the argument of Forde and Zhang (2017) where the authors carry along a factor $\hat{\varepsilon}/\varepsilon = \varepsilon^{2H-1}$. (This provides a *unified view on rough and moderate deviations*.) The remaining details then follow essentially 'Appendix C. Proof of Corollary 4.13., part (ii) upper bound' of Forde and Zhang (2017), noting perhaps that the authors use their assumptions to show validity of what we simply assumed as condition (iiib), and also that one works with the quadratic rate function $I''(0)x^2 = x^2/2\sigma_0^2$ throughout. ■

REMARK 7.2 By an easy argument similar to 'Appendix C. Proof of Corollary 4.13., part (i) lower bound' of Forde and Zhang (2017) one sees that validity of the call price upper bound (iiic) implies the corresponding digital call price upper bound (40.) For this reason, we only emphasized (iiic) but not (40) in Section 2.

In a classical work Azencott (1982) (see also Azencott 1985; Ben Arous 1988, Théorème 2) obtained asymptotic expansions of functionals of Laplace type on Wiener space, of the type '$E[\exp(-F(X^\varepsilon)/\varepsilon^2)]$', for small noise diffusions X^ε. This refines the large deviation (equivalently: Laplace) principle of Freidlin–Wentzell for small noise diffusions. In a nutshell, for fixed $X_0 = x$, Azencott gets expansions of the form $\mathrm{e}^{-c/\varepsilon^2}(\alpha_0 + \alpha_1\varepsilon\cdots)$. His ideas (used by virtually all subsequent works in this direction) are a Girsanov transform, to make the minimizing path 'typical', followed by localization around the minimizer (justified by a good large deviation principle), and finally a local (stochastic Taylor) type analysis near the minimizer. None of these ingredients rely on the Markovian structure (or, relatedly, PDE arguments). As a consequence (and motivation for this work) such expansions were also obtained in the (non-Markovian) context of rough differential equations driven by fractional Brownian motion (Inahama 2013; Baudoin and Ouyang 2015) with $H < 1/2$.

And yet, our situation is different in the sense that call price Wiener functionals do not fit the form studied by Azencott and others, nor can we in fact expect a similar expansion: Example 3.3 gives a Black-Scholes call price expansion of the form constant times $\mathrm{e}^{-c\varepsilon^2}(\varepsilon^3 + \cdots)$. Azencott's ideas are nonetheless very relevant to us: we already used the Girsanov formula

in Theorem 3.2 in order to have a tractable expression for J. It thus 'only' remains to carry out the localization and do some local analysis.

PROPOSITION 7.3 *Let $x>0$ and $\beta \in (0,H)$. Then the factor J is negligible in the sense that, for every $\theta>0$,*

$$\varepsilon^{\theta} \log J(\varepsilon, x\varepsilon^{2\beta}) \to 0 \quad \text{as } \varepsilon \to 0.$$

Proof Step 1. Localization Write $x_\varepsilon := x\varepsilon^{2\beta}, \hat{x}_\varepsilon := x_\varepsilon \varepsilon^{1-2H} = x\varepsilon^{1-2H+2\beta}$. By definition,

$$E[(\mathrm{e}^{X_1^\varepsilon} - \mathrm{e}^{\hat{x}_\varepsilon})^+]\, \mathrm{e}^{I(x_\varepsilon)/\hat{\varepsilon}^2}\, \mathrm{e}^{-\hat{x}_\varepsilon} = J(\varepsilon, x_\varepsilon).$$

Fix $x, \delta > 0$ and write $\delta_\varepsilon = \delta\varepsilon^{2\beta}$. We claim that (the positive quantity)

$$J(\varepsilon, x_\varepsilon) - J_{\delta_\varepsilon}(\varepsilon, x_\varepsilon) = \mathrm{e}^{I(x_\varepsilon)/\hat{\varepsilon}^2}\, \mathrm{e}^{-\hat{x}_\varepsilon} E[(\mathrm{e}^{X_1^\varepsilon} - \mathrm{e}^{\hat{x}_\varepsilon}) 1_{\hat{X}_1^\varepsilon > x_\varepsilon + \delta_\varepsilon}] \tag{41}$$

is exponentially small, in the sense that, for some $c>0$ and $\bar{\varepsilon}^2 = \varepsilon^{4H-4\beta}$,

$$J(\varepsilon, x_\varepsilon) - J_{\delta_\varepsilon}(\varepsilon, x_\varepsilon) = O\left(\mathrm{e}^{-c/\bar{\varepsilon}^2}\right).$$

There is a battle here between the exploding factor $\mathrm{e}^{I(x_\varepsilon)/\hat{\varepsilon}^2}$, with exponent

$$\frac{I(x_\varepsilon)}{\hat{\varepsilon}^2} \sim \frac{I''(0)\,(x_\varepsilon)^2}{2\hat{\varepsilon}^2} = \frac{I''(0)x^2}{2\varepsilon^{4H-4\beta}},$$

and on the other hand

$$E[(\mathrm{e}^{X_1^\varepsilon} - \mathrm{e}^{\hat{x}_\varepsilon}) 1_{\hat{X}_1^\varepsilon > x_\varepsilon + \delta_\varepsilon}] \le \exp\left(-\frac{(x+\delta)^2 + o(1)}{2\sigma_0^2 \varepsilon^{4H-4\beta}}\right)$$

where the given estimate is an easy consequence of Lemma 7.1. Since $I''(0) = 1/\sigma_0^2$ we see that the last factor 'exponentially over-compensates' the rest, so that the difference is indeed exponentially negligible.

Step 2. Upper bound. For any $x>0$, recall that $\hat{U}^{\varepsilon,x} = \hat{U}^\varepsilon$ decomposes into a Gaussian random variable $g_1 = g_1^x$ and remainder $R_2^{\varepsilon,x} = R_2^\varepsilon$. In order to control this remainder *without* imposing boundedness assumption on $\sigma(.)$, we will crucially used a 'localized remainder tail estimate' as given in Proposition 7.4 below. We have, for any $\varepsilon \in (0,1]$,

$$\begin{aligned} J_\delta(\varepsilon, x) &= E\left[\mathrm{e}^{-(I'(x)/\hat{\varepsilon}^2)\hat{U}^\varepsilon}\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{U}^\varepsilon\right) - 1\right)\mathrm{e}^{I'(x)R_2^\varepsilon} \mathbf{1}_{\hat{U}^\varepsilon \in [0,\delta_\varepsilon]}\right] \\ &\le (\mathrm{e}^\delta - 1) E[\mathrm{e}^{-I'(x)/\hat{\varepsilon} g_1^x}; \hat{U}^{\varepsilon,x} \in [0,\delta]]. \end{aligned} \tag{42}$$

To proceed, recall $\hat{\varepsilon}^{-1} g_1^x = \hat{\varepsilon}^{-2}\hat{U}^{\varepsilon,x} - R_2^{\varepsilon,x}$ so that, for any $\kappa > 0$,

$$\begin{aligned} \mathrm{e}^{-(I'(x)/\hat{\varepsilon})g_1^x} &= \mathrm{e}^{-(I'(x)/\hat{\varepsilon})g_1^x} \mathbf{1}_{\left|\hat{\varepsilon}\hat{B}\right|_{\infty;[0,1]} \ge \kappa} \\ &\quad + \mathrm{e}^{-(I'(x)/\hat{\varepsilon}^2)\hat{U}^{\varepsilon,x}}\, \mathrm{e}^{I'(x)R_2^{\varepsilon,x}} \mathbf{1}_{\left|\hat{\varepsilon}\hat{B}\right|_{\infty;[0,1]} < \kappa}. \end{aligned}$$

Since $I'(x) > 0$ for small enough $x>0$, it follows that $-(I'(x)/\hat{\varepsilon}^2)\hat{U}^{\varepsilon,x} < 0$ on the event $\{\hat{U}^{\varepsilon,x} \in [0,\delta]\}$, which leads us to

$$\begin{aligned} J_\delta(\varepsilon, x) &\le (\mathrm{e}^\delta - 1) E[\mathrm{e}^{-(I'(x)/\hat{\varepsilon})g_1^x}; |\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} \ge \kappa] \\ &\quad + (\mathrm{e}^\delta - 1) E[\mathrm{e}^{I'(x)R_2^{\varepsilon,x}}; |\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} < \kappa] \\ &\le (\mathrm{e}^\delta - 1)\sqrt{E[\mathrm{e}^{-(2I'(x)/\hat{\varepsilon})g_1^x}]}\sqrt{P\left[|\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} \ge \kappa\right]} \\ &\quad + (\mathrm{e}^\delta - 1) C \end{aligned}$$

where, by Proposition 7.4, the constant $C = C(\kappa)$ is uniform in small ε and x. The square-root terms are computed resp. (Fernique) estimated by

$$\exp\left(\frac{(I'(x))^2 \mathbb{V}(g_1^x)}{\hat{\varepsilon}^2}\right) \times \exp(-c\kappa^2/\hat{\varepsilon}^2)$$

for some $c>0$ which depends on the law of B (hence H), but is uniform in ε and x. Hence, for x small enough, the resulting exponent $(I'(x))^2\mathbb{V}(g_1^x) - c\kappa^2$ is negative, which is more than enough to conclude the upper bound.

Step 3. Lower bound. Write $E_{\delta,\kappa}[\cdot] = E[\cdot \mathbf{1}_{\hat{U}^{\varepsilon,x} \in [0,\delta_\varepsilon]} \mathbf{1}_{|\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} < \kappa}]$ and estimate

$$\begin{aligned} &E_{\delta,\kappa}\left[\mathrm{e}^{-(I'(x)/\hat{\varepsilon}^2)\hat{U}^\varepsilon/2}\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{U}^\varepsilon\right) - 1\right)^{1/2}\right] \\ &\quad = E_{\delta,\kappa}\left[\mathrm{e}^{-(I'(x)/\hat{\varepsilon}^2)\hat{U}^\varepsilon/2}\left(\exp\left(\frac{\varepsilon}{\hat{\varepsilon}}\hat{U}^\varepsilon\right) - 1\right)^{1/2} \mathrm{e}^{I'(x)R_2^\varepsilon/2} \right. \\ &\quad \left. \times\, \mathrm{e}^{-I'(x)R_2^\varepsilon/2}\right] \\ &\quad \le J_\delta(\varepsilon, x)^{1/2}\, E_{\delta,\kappa}\left[\mathrm{e}^{-I'(x)R_2^\varepsilon}\right]^{1/2} \end{aligned}$$

where we used Cauchy–Schwarz and discarded the event $\{|\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} < \kappa\}$. The localized remainder estimate provides an upper bound on $E_{\delta,\kappa}[\mathrm{e}^{-I'(x)R_2^\varepsilon}]$, uniformly over small (enough) ε and x.

It then suffices to get a suitable lower bound of the left-hand side above. Indeed, for $u \in [0, \hat{\varepsilon}^2\eta] = [0, \varepsilon^{4H}\eta]$, with η small enough, not dependent on ε,

$$u \mapsto (\mathrm{e}^{(\varepsilon/\hat{\varepsilon})u} - 1)^{1/2}\, \mathrm{e}^{-(I'(x)/\hat{\varepsilon}^2)u/2} \ge \gamma\left(\frac{\varepsilon}{\hat{\varepsilon}}u\right)^{1/2} \tag{43}$$

for a constant $\gamma > 0$ which can also be taken uniformly in small x, ε. Then estimate

$$\begin{aligned} &E_{\delta,\kappa}[(\mathrm{e}^{(\varepsilon/\hat{\varepsilon})\hat{U}^\varepsilon} - 1)^{1/2}\, \mathrm{e}^{-(I'(x)/2\varepsilon^2)\hat{U}^\varepsilon}] \\ &\quad \ge \gamma\varepsilon^{1/2-H} E[|\hat{U}^\varepsilon|^{1/2} 1_{\hat{U}^\varepsilon \in [0,\hat{\varepsilon}^2\eta]} 1_{|\varepsilon B|_{\infty;[0,1]} < \kappa}]. \end{aligned}$$

As a quick sanity check, pretend zero remainder so that $\hat{U}^\varepsilon = \hat{\varepsilon} g_1$: dropping further the (exponentially close to probability one) event $\{|\varepsilon B|_{\infty;[0,1]} < \kappa\}$, a Gaussian computation then

shows that we are left with ($\gamma\varepsilon^{1/2-H}$ times $\hat{\varepsilon}^{1/2}$ times)

$$E[|g_1|^{1/2}; g_1 \in [0,\hat{\varepsilon}]] \sim (const)\hat{\varepsilon}^{3/2}.$$

In general, set $V^\varepsilon = \hat{U}^\varepsilon/\hat{\varepsilon} = g_1 + \hat{\varepsilon}R_2 s^\varepsilon$, so that†

$$E_\kappa\left[|\hat{U}^\varepsilon|^{1/2}; \hat{U}^\varepsilon \in [0,\hat{\varepsilon}^2\eta]\right] = \hat{\varepsilon}^{1/2}\, E_\kappa[|V^\varepsilon|^{1/2}; V^\varepsilon \in [0,\hat{\varepsilon}\eta]].$$

At this stage, it is difficult to treat $\hat{\varepsilon}R^\varepsilon$ as perturbation of g since, on the given event $\{V^\varepsilon \in [0,\hat{\varepsilon}\eta]\}$, all terms are of order $\hat{\varepsilon}$. We can solve this issue by realizing that we can replace, throughout, x by $x_\varepsilon = x\varepsilon^{2\beta}$. Since $I'(x_\varepsilon) \sim (const)x_\varepsilon$, with see from (43), that in the above estimate the event $\hat{U}^\varepsilon \in [0,\hat{\varepsilon}^2\eta] = [0,\varepsilon^{4H}\eta]$ (resp. $V^\varepsilon \in [0,\hat{\varepsilon}\eta] = [0,\varepsilon^{2H}\eta]$) can be replaced by $\hat{U}^\varepsilon \in [0,\varepsilon^{4H-2\beta}\eta]$ (resp. $V^\varepsilon \in [0,\varepsilon^{2H-2\beta}\eta]$), possibly with an insignificantly modified constant η. It is now straight-forward to show that the behavior of $E_\kappa[|V^\varepsilon|^{1/2}; V^\varepsilon \in [0,\varepsilon^{2H-2\beta}\eta]]$ is of the same order as $E[g^{1/2}; g \in [0,\varepsilon^{2H-2\beta}\eta]]$, the correct behavior (i.e. positive power of ε) is obtained by spelling out the (Gaussian) integral. ■

PROPOSITION 7.4 Localized remainder tail estimate *For every $\kappa > 0$, there exists $c_1, c_2 > 0$ such that, for all r and uniformly in small ε, x we have*

$$P\left[\left|R_2^\varepsilon\right| > r, |\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} < \kappa\right] \le c_1 \exp(-c_2 r)$$

Proof We decompose $\hat{\varepsilon}^2 R_2^\varepsilon = M^\varepsilon + N^\varepsilon$ in terms of the (local) martingale

$$M^\varepsilon := \hat{\varepsilon}\int_0 \left[\sigma\left(\hat{\varepsilon}\hat{B} + \hat{f}\right) - \sigma\left(\hat{f}\right)\right] \mathrm{d}[\bar{\rho}W + \rho B]$$

and the (bounded variation) process

$$N^\varepsilon := \int_0 \left[\sigma\left(\hat{\varepsilon}\hat{B} + \hat{f}\right) - \sigma\left(\hat{f}\right) - \sigma'\left(\hat{f}\right)\hat{\varepsilon}B\right] \mathrm{d}[\bar{\rho}h + \rho f]$$
$$- \frac{1}{2}\varepsilon\hat{\varepsilon}\int_0 \sigma^2\left(\hat{\varepsilon}\hat{B} + \hat{f}\right)\mathrm{d}t.$$

Let $\tau^{\varepsilon,\kappa}$ be the stopping time when $\hat{\varepsilon}\hat{B}$ first leaves the uniform ball of radius κ. Then

$$M_t^{\kappa,\varepsilon} := M_{t\wedge\tau^{\varepsilon,\kappa}}^\varepsilon$$

still yields a (local) martingale. The point is that $\{|\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} < \kappa\} = \{\tau^{\varepsilon,\kappa} > 1\}$. On this event, $M^\varepsilon|_{[0,1]} = M^{\kappa,\varepsilon}|_{[0,1]}$ and we can thus replace M^ε, in the definition of the remainder, by $M^{\kappa,\varepsilon}$. Let $K = K^{\kappa,x}$ be the κ-fattening of $\{f(t) : 0 \le t \le 1\}$, recall $f = f^x$, then, for $t \in [0,1]$,

$$\mathrm{d}[M^{\kappa,\varepsilon}]_t/\mathrm{d}t = \hat{\varepsilon}^2(\sigma(\hat{\varepsilon}\hat{B}_t + \hat{f}_t) - \sigma(f_t))^2 \le \hat{\varepsilon}^4 \left\|\sigma'\right\|_{\infty;K}^2 |\hat{B}_t|^2.$$

Clearly, we can replace K by $\tilde{K}^\kappa$ which contains all $K^{\kappa,x}$ for small x. To summarize, we have, on the event $\{|\hat{\varepsilon}\hat{B}|_{\infty;[0,1]} < \kappa\}$,

$$R^\varepsilon(\cdot) = \hat{\varepsilon}^{-2}M^{\kappa,\varepsilon} + \hat{\varepsilon}^{-2}N^\varepsilon$$

with $[\hat{\varepsilon}^{-2}M^{\kappa,\varepsilon}] = O(|\hat{B}|_{\infty;[0,1]}^2)$ and, as seen by a similar (but easier) reasoning, $\hat{\varepsilon}^{-2}N^\varepsilon = O(|\hat{B}|_{\infty;[0,1]}^2)$, always for fixed $\kappa > 0$, but uniformly in small ε (equivalently, $\hat{\varepsilon}$) and small $x > 0$. This clearly shows that $\hat{\varepsilon}^{-2}N^\varepsilon$ has exponential tails. The same is true for the martingale part, whose bracket is $O(\text{Gaussian}^2)$. This is exactly the situation for the 'model' martingal increment $2\int_0^1 B\,\mathrm{d}B = B_1^2 - 1$ which clearly has exponential tails. To make this rigorous, recall that Gaussian resp. exponential tails are characertized by $O(\sqrt{p})$ resp. $O(p)$-growth of the L^p-norms. The statement is then an easy consequence of the sharp (upper) BDG constant (Carlen and Kree 1991), known to be $O(\sqrt{p})$. ■

† Write E_κ for the expected valued restricted to the event $\{|\varepsilon B|_{\infty;[0,1]} < \kappa\}$

8. Proof of the implied volatility expansion

With Theorem 3.2 in place, we now turn to the proof of the implied volatility expansion, formulated in Theorem 3.6.

Proof of Theorem 3.6 We will use an asymptotic formula for the dimensionless implied variance

$$V_t^2 = t\sigma_{\text{impl}}(k_t,t)^2, \quad t > 0,$$

obtained in Gao and Lee (2014). It follows from the first formula in Remark 7.3 in Gao and Lee (2014) that

$$V_t^2 - \frac{k_t^2}{2L_t} = \mathcal{O}\left(\frac{k_t^2}{L_t^2}(k_t + |\log k_t| + \log L_t)\right), \quad t \to 0, \tag{44}$$

where $L_t = -\log c(k_t,t)$, $t > 0$.

We will need the following formula that was established in the proof of Theorem 3.4:

$$L_t = \frac{I(kt^\beta)}{t^{2H}} + \mathcal{O}(t^{-\theta}) \tag{45}$$

as $t \to 0$, for all $x \ge 0$ and $\beta \in [0,H)$ and any $\theta > 0$. Let us first assume $2H/(n+1) \le \beta < 2H/n$. Using the energy expansion, we obtain from (45) that

$$L_t = \sum_{i=2}^n \frac{I^{(i)}(0)}{i!}k^i t^{i\beta-2H} + \mathcal{O}\left(t^{-\theta}\right) = \frac{I''(0)}{2}k^2 t^{2\beta-2H}$$
$$\times\left[1 + \sum_{i=3}^n \frac{2I^{(i)}(0)}{i!I''(0)}k^{i-2}t^{(i-2)\beta} + \mathcal{O}\left(t^{2H-2\beta-\theta}\right)\right] \tag{46}$$

as $t \to 0$. The second term in the brackets on the right-hand side of (46) disappears if $n = 2$.

REMARK 8.1 Suppose $n \ge 2$ and $2H/(n+1) \le \beta < 2H/n$. Then formula (46) is optimal. Next, suppose $n \ge 2$ and $0 < \beta < 2H/(n+1)$. In this case, there exists $m \ge n+1$ such that $2H/(m+1) \le \beta < 2H/m$, and hence (46) holds with m instead of n. However, we can replace m by n, by making the error term worse. It is not hard to see that the following

formula holds for all $n \geq 2$ and $0 < \beta < 2H/(n+1)$:

$$L_t = \sum_{i=2}^{n} \frac{I^{(i)}(0)}{i!} k^i t^{i\beta-2H} + \mathcal{O}\left(t^{(n+1)\beta-2H}\right) = \frac{I''(0)}{2} k^2 t^{2\beta-2H} \times \left[1 + \sum_{i=3}^{n} \frac{2I^{(i)}(0)}{i!I''(0)} k^{i-2} t^{(i-2)\beta} + \mathcal{O}\left(t^{(n-1)\beta}\right)\right] \tag{47}$$

as $t \to 0$ provided we choose θ small enough.

Let us continue the proof of Theorem 3.6. Since $k_t \approx t^{1/2-H+\beta}$ and $L_t \approx t^{2\beta-2H}$ as $t \to 0$, (44) implies that

$$V_t^2 = \frac{k^2 t^{1-2H+2\beta}}{2L_t} + \mathcal{O}\left(t^{1+2H-2\beta-\theta}\right), \quad t \to 0. \tag{48}$$

Next, using the Taylor formula for the function $u \mapsto 1/(1+u)$, and setting

$$u = \sum_{i=3}^{n} \frac{2I^{(i)}(0)}{i!I''(0)} k^{i-2} t^{(i-2)\beta} + \mathcal{O}(t^{2H-2\beta-\theta}),$$

we obtain from (46) that

$$(2L_t)^{-1} = \frac{t^{2H-2\beta}}{k^2 I''(0)} \left[\sum_{j=0}^{n-2} (-1)^j u^j + \mathcal{O}(u^{n-1})\right]$$

as $t \to 0$. It follows from $2H/(n+1) \leq \beta < 2H/n$ that $(n-1)\beta \geq 2H - 2\beta$, and hence

$$\begin{aligned}(2L_t)^{-1} &= \frac{t^{2H-2\beta}}{k^2 I''(0)} \left[\sum_{j=0}^{n-2} (-1)^j u^j\right] + \mathcal{O}(t^{4H-4\beta-\theta}) \\ &= \frac{t^{2H-2\beta}}{k^2 I''(0)} \left[\sum_{j=0}^{n-2} (-1)^j \left(\sum_{i=3}^{n} \frac{2I^{(i)}(0)}{i!I''(0)} k^{i-2} t^{(i-2)\beta}\right)^j\right] \\ &\quad + \mathcal{O}(t^{4H-4\beta-\theta})\end{aligned}$$

as $t \to 0$. Now, (48) gives

$$V_t^2 = \frac{t}{I''(0)} \left[\sum_{j=0}^{n-2} (-1)^j \left(\sum_{i=3}^{n} \frac{2I^{(i)}(0)}{i!I''(0)} k^{i-2} t^{(i-2)\beta}\right)^j\right] + \mathcal{O}\left(t^{1+2H-2\beta-\theta}\right)$$

as $t \to 0$. Finally, by canceling a factor of t in the previous formula, we obtain formula (14) for $2H/(n+1) \leq \beta < 2H/n$. The proof in the case where $\beta \leq 2H/(n+1)$ is similar. Here we take into account Remark 8.1. This completes the proof of Theorem 3.6. ■

Acknowledgments

Two referees are thanked for their useful comments. We further thank Martin Forde for valuable feedback.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

We gratefully acknowledge financial support through DFG research grants FR2943/2 and BA5484/1 (C. Bayer, P.K. Friz, B. Stemper), European Research Council Grant CoG-683164 (P.K. Friz), and SNF Early Postdoc Mobility Grant 165248 (B. Horvath) respectively.

References

Alòs, E., León, J.A. and Vives, J., On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance Stoch.*, 2007, **11**(4), 571–589.

Azencott, R., Formule de Taylor stochastique et développement asymptotique d'intégrales de Feynman. In *Seminar on Probability, XVI, Supplement*, volume 921 of *Lecture Notes in Math.*, pp. 237–285, 1982 (Springer: Berlin-New York).

Azencott, R., Petites perturbations aléatoires des systemes dynamiques: développements asymptotiques. *Bull. Sci. Math.*, 1985, **109**(3), 253–308.

Baudoin, F. and Ouyang, C., On small time asymptotics for rough differential equations driven by fractional Brownian motions. In *Large Deviations and Asymptotic Methods in Finance*, edited by P.K. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, and J. Teichmann, pp. 413–438, 2015 (Springer International Publishing: Cham).

Bayer, C., Friz, P.K., Gassiat, P., Martin, J. and Stemper, B., A regularity structure for rough volatility. Preprint, 2017. arXiv:1710.07481.

Bayer, C., Friz, P.K. and Gatheral, J., Pricing under rough volatility. *Quant. Finance*, 2016, **16**(6), 887–904.

Ben Arous, G., Methods de Laplace et de la phase stationnaire sur l'espace de Wiener. *Stochastics*, 1988, **25**(3), 125–153.

Bennedsen, M., Lunde, A. and Pakkanen, M.S., Decoupling the short- and long-term behavior of stochastic volatility. Preprint, 2016. arXiv:1610.00332.

Bennedsen, M., Lunde, A. and Pakkanen, M.S., Hybrid scheme for Brownian semistationary processes. *Finance Stoch.*, 2017, **21**(4), 931–965.

Bismut, J.-M., *Large Deviations and the Malliavin Calculus*. Progress in Mathematics, Vol. 45. 1984 (Birkhäuser Boston, Inc.: Boston, MA).

Carlen, E. and Kree, P., Estimates on iterated stochastic integrals. *Ann. Probab.*, 1991, **19**(1), 354–368.

Cass, T. and Friz, P., Densities for rough differential equations under Hörmander's condition. *Ann. Math.*, 2010, **0**, 2115–2141.

Decreusefond, L., Stochastic integration with respect to Volterra processes. *Ann. l. H. Poincare Probab. Statist.*, 2005, **41**(2), 123–149.

Deuschel, J.-D., Friz, P.K., Jacquier, A. and Violante, S., Marginal density expansions for diffusions and stochastic volatility I: Theoretical foundations. *Comm. Pure Appl. Math.*, 2014a, **67**(1), 40–82.

Deuschel, J.-D., Friz, P.K., Jacquier, A. and Violante, S., Marginal density expansions for diffusions and stochastic volatility II: Applications. *Comm. Pure Appl. Math.*, 2014b, **67**(2), 321–350.

Deuschel, J.-D. and Stroock, D.W., *Large Deviations, Vol. 137*, 1989 (Academic Press: Boston, MA).

El Euch, O. and Rosenbaum, M., The characteristic function of rough Heston models. Preprint, 2016. To appear in *Math. Finance*.

Forde, M. and Jacquier, A., Small-time asymptotics for implied volatility under the heston model. *Int. J. Theoret. Appl. Finance*, 2009, **12**(06), 861–876.

Forde, M. and Zhang, H., Asymptotics for rough stochastic volatility models. *SIAM J. Financ. Math.*, 2017, **8**(1), 114–145.
Friz, P.K. and Gassiat, P., Martingality and moments for lognormal rough volatility. In preparation, 2018.
Friz, P.K., Gerhold, S. and Pinter, A., Option Pricing in the Moderate Deviations Regime. *Math. Finance*, 2018, **28**(3), 962–988.
Friz, P. and Hairer, M., *A Course on Rough Paths*, 2014 (Springer: Cham).
Fukasawa, M., Asymptotic analysis for stochastic volatility: martingale expansion. *Finance Stoch.*, 2011, **15**(4), 635–654.
Fukasawa, M., Short-time at-the-money skew and rough fractional volatility. *Quant. Finance*, 2017, **17**(2), 189–198.
Gao, K. and Lee, R., Asymptotics of implied volatility to arbitrary order. *Finance Stoch.*, 2014, **18**(2), 349–392.
Gatheral, J., *The Volatility Surface: A Practitioner's Guide*, 2011 (John Wiley & Sons: Hoboken, NJ).
Gatheral, J., Jaisson, T. and Rosenbaum, M., Volatility is rough. Preprint, 2014. To appear in *Quant. Finance*.
Guennoun, H., Jacquier, A. and Roome, P., Asymptotic behaviour of the fractional Heston model. Preprint, 2014. arXiv:1411.7653.
Gulisashvili, A., Large deviation principle for Volterra type fractional stochastic volatility models. *ArXiv e-prints*, October 2017. To appear in *SIAM J. Financ. Math*.
Inahama, Y., Laplace approximation for rough differential equation driven by fractional brownian motion. *Ann. Probab.*, 2013, **41**(1), 170–205.
Jacquier, A., Pakkanen, M.S. and Stone, H., Pathwise large deviations for the Rough Bergomi model. *ArXiv e-prints*, June 2017.
Jourdain, B., Loss of martingality in asset price models with lognormal stochastic volatility. *Int. J. Theoret. Appl. Finance*, 2004, **13**, 767–787.
Lamperti, J., Semi-stable stochastic processes. *Trans. Am. Math. Soc.*, 1962, **104**(1), 62–78.
Lions, P.-L. and Musiela, M., Correlations and bounds for stochastic volatility models. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 2007, **24**(1), 1–16.
Medvedev, A. and Scaillet, O., A simple calibration procedure of stochastic volatility models with jumps by short term asymptotics. Preprint, 2003. Available at SSRN 477441.
Medvedev, A. and Scaillet, O., Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Rev. Financ. Stud.*, 2007, **20**(2), 427–459.
Mijatović, A. and Tankov, P., A new look at short-term implied volatility in asset price models with jumps. *Math. Finance*, 2016, **26**(1), 149–183.
Muhle-Karbe, J. and Nutz, M., Small-time asymptotics of option prices and first absolute moments. *J. Appl. Probab.*, 2011, **48**(4), 1003–1020.
Olver, F.W.J., Lozier, D.W., Boisvert, R.F. and Clark, C.W (Eds.), *NIST Handbook of Mathematical Functions*, 2010 (Cambridge University Press: New York).
Osajima, Y., The asymptotic expansion formula of implied volatility for dynamic SABR model and FX hybrid model. Preprint, 2007. Available at SSRN 965265.
Osajima, Y., General asymptotics of Wiener functionals and application to implied volatilities, In *Large Deviations and Asymptotic Methods in Finance*, edited by P.K. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, and J. Teichmann, pp. 137–173, 2015 (Springer International Publishing: Cham).
Pham, H., Large deviations in mathematical finance, 2010. Available online at: https://www.lpsm.paris/pageperso/pham/GD-finance.pdf.
Sin, C.A., Complications with stochastic volatility models. *Adv. Appl. Probab.*, 1998, **30**(1), 256–268.

Appendix. Auxiliary lemmas

In this section we provide and prove some auxiliary lemmas, which are used in the preparations to the proof of Theorem 3.2. We start with a technical Lemma, that justifies the derivation.

LEMMA A.1 *Assume* $\sigma(.) > 0$ *and* $|\rho| < 1$. *Then* $\mathcal{K}^x$ *is a Hilbert manifold near any* $\mathfrak{h} := (h,f) \in \mathcal{K}^x \subset \mathfrak{H} := H_0^1 \times H_0^1$.

Proof Similar to Bismut (1984, p. 25) we need to show that $D\varphi_1(\mathfrak{h})$ is surjective where $\varphi_1(\mathfrak{h}) : \mathfrak{H} \to \mathbb{R}$ with

$$\varphi_1(\mathfrak{h}) = \varphi_1(h,f) = \int_0^1 \sigma(\hat{f})\, \mathrm{d}\left(\bar{\rho}h + \rho f\right).$$

From

$$\begin{aligned}\varphi_1\left(\mathfrak{h} + \delta\mathfrak{h}'\right) &= \int_0^1 \sigma(\hat{f} + \delta\hat{f}')\, \mathrm{d}\left(\bar{\rho}h + \rho f + \delta(\bar{\rho}h' + \rho f')\right) \\ &= \varphi_1(\mathfrak{h}) + \delta \int_0^1 \sigma(\hat{f})\, \mathrm{d}(\bar{\rho}h' + \rho f') \\ &\quad + \delta \int_0^1 \sigma'(\hat{f})\hat{f}'\, \mathrm{d}\left(\bar{\rho}h + \rho f\right) + o(\delta).\end{aligned}$$

the functional derivative $D\varphi_1(\mathfrak{h})$ can be computed explicitly. In fact, even the computation

$$\left(D\varphi_1(\mathfrak{h}), (h',0)\right) = \bar{\rho} \int_0^1 \sigma(\hat{f})\, \mathrm{d}h'$$

is sufficient to guarantee surjectivity of $D\varphi_1(\mathfrak{h})$. ■

We now give the proof of Lemma 6.3, which determines the form of the Girsanov measure change (36) for the minimizing configuration.

LEMMA A.2 (i) *Any optimal control* $\mathfrak{h}^0 = (h^x, f^x) \in \mathcal{K}^x$ *is a critical point of*

$$\mathfrak{h} = (h,f) \mapsto -I\left(\varphi_1^{\mathfrak{h}}\right) + \tfrac{1}{2}\|\mathfrak{h}\|_{\mathfrak{H}}^2;$$

(ii) *it holds that*

$$\int_0^1 \dot{h}^x\, \mathrm{d}W + \int_0^1 \dot{f}^x\, \mathrm{d}B = I'(x)\, g_1.$$

Proof (Step 1) Write $\mathfrak{h} = (h,f)$ and

$$\varphi_1(\mathfrak{h}) = \varphi_1(h,f) = \int_0^1 \sigma(\hat{f})\, \mathrm{d}\left(\bar{\rho}h + \rho f\right).$$

Let $\mathfrak{h}^0 = (h^x, f^x) \in \mathcal{K}^x$ an optimal control. Then

$$\mathfrak{Ker} D\varphi_1\left(\mathfrak{h}^0\right) = T_{\mathfrak{h}^0}\mathcal{K}^x = \left\{\mathfrak{h} \in \mathfrak{H}^1 : D\varphi_1(\mathfrak{h}) = 0\right\}.$$

(This requires $\mathcal{K}^x$ to be a Hilbert manifold near $\mathfrak{h}^0$, as was seen in the last lemma.)

(Step 2) For fixed $\mathfrak{h} \in \mathfrak{H}$, define

$$u(t) := -I\left(\varphi_1^{\mathfrak{h}^0 + t\mathfrak{h}}\right) + \tfrac{1}{2}\left\|\mathfrak{h}^0 + t\mathfrak{h}\right\|_{\mathfrak{H}}^2 \geq 0$$

with equality at $t = 0$ (since $x = \varphi_1^{\mathfrak{h}^0}$ and $I(x) = \frac{1}{2}\|\mathfrak{h}^0\|_{\mathfrak{H}}^2$) and non-negativity for all t because $\mathfrak{h}^0 + t\mathfrak{h}$ is an admissible control for reaching $\tilde{x} = \varphi_1^{\mathfrak{h}^0 + t\mathfrak{h}}$ (so that $I(\tilde{x}) = \inf\{\cdots\} \leq \frac{1}{2}\|\mathfrak{h}^0 + t\mathfrak{h}\|_{\mathfrak{H}}^2$.)

(Step 3) We note that $\dot{u}(0) = 0$ is a consequence of $u \in C^1$ near 0, $u(0) = 0$ and $u \geq 0$. In other words, $\mathfrak{h}^0$ is a critical point for

$$\mathfrak{H}^1 \ni \mathfrak{h} \mapsto -I\left(\varphi_1^{\mathfrak{h}}\right) + \tfrac{1}{2}\|\mathfrak{h}\|_{\mathfrak{H}}^2.$$

(Step 4) The functional derivative of this map at $\mathfrak{h}^0$ must hence be zero. In particular, for all $\mathfrak{h} \in \mathfrak{H}$,

$$0 \equiv -I'\left(\varphi_1^{\mathfrak{h}^0}\right)\left\langle D\varphi_1\left(\mathfrak{h}^0\right),\mathfrak{h}\right\rangle + \left\langle \mathfrak{h}^0,\mathfrak{h}\right\rangle$$
$$= -I'(x)\left\langle D\varphi_1\left(\mathfrak{h}^0\right),\mathfrak{h}\right\rangle + \left\langle \mathfrak{h}^0,\mathfrak{h}\right\rangle.$$

(Step 5) With $\mathfrak{h}^0 = (h^x, f^x)$ and $\mathfrak{h} = (h, f)$

$$\left\langle D\varphi_1\left(\mathfrak{h}^0\right),\mathfrak{h}\right\rangle = \left.\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0}\int_0^1 \sigma(\hat{f}^x + \varepsilon\hat{f})\,\mathrm{d}$$
$$\times\left(\bar{\rho}h^x + \rho f^x + \varepsilon(\bar{\rho}h + \rho f)\right)$$
$$= \int_0^1 \sigma(\hat{f}^x)\,\mathrm{d}(\bar{\rho}h + \rho f) + \int_0^1 \sigma'(\hat{f}^x)\hat{f}\,\mathrm{d}\left(\bar{\rho}h^x + \rho f^x\right)$$

By continuous extension, replace $\mathfrak{h} = (h, f)$ by (W, B) above and note that

$$\left\langle D\varphi_1\left(\mathfrak{h}^0\right),(W,B)\right\rangle = g_1$$

since indeed $g_1 = \int_0^1 \sigma(\hat{f}_t)\,\mathrm{d}(\bar{\rho}W_t + \rho B_t) + \sigma'(\hat{f}_t)\hat{B}_t\,\mathrm{d}(\bar{\rho}h_t + \rho f_t)$. Hence

$$\int_0^1 \dot{h}^x\,\mathrm{d}W + \int_0^1 \dot{f}^x\,\mathrm{d}B = I'(x)\,g_1.$$

■