Local volatility under rough volatility

F. Bourgey[*1,2], S. De Marco[†1], P. K. Friz[‡3], and P. Pigato [§4]

[1]Centre de Mathématiques Appliquées (CMAP), CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, France
[2]Bloomberg L.P., Quantitative Research, London, UK
[3]Technische Universität Berlin and Weierstraß-Institut, Berlin, Germany
[4]Department of Economics and Finance, Università Roma Tor Vergata, Rome, Italy

April 6, 2022

Abstract

Several asymptotic results for the implied volatility generated by a rough volatility model have been obtained in recent years (notably in the small-maturity regime), providing a better understanding of the shapes of the volatility surface induced by rough volatility models, and supporting their calibration power to S&P500 option data. Rough volatility models also generate a local volatility surface, via the so-called Markovian projection of the stochastic volatility. We complement the existing results on the implied volatility by studying the asymptotic behavior of the local volatility surface generated by a class of rough stochastic volatility models, encompassing the rough Bergomi model. Notably, we observe that the celebrated "1/2 skew rule" linking the short-term at-the-money skew of the implied volatility to the short-term at-the-money skew of the local volatility, a consequence of the celebrated "harmonic mean formula" of [Berestycki, Busca, and Florent, QF 2002], is replaced by a new rule: the ratio of the at-the-money implied and local volatility skews tends to the constant $1/(H + 3/2)$ (as opposed to the constant $1/2$), where H is the regularity index of the underlying instantaneous volatility process.

Contents

---

[*]florian.bourgey@polytechnique.edu
[†]stefano.de-marco@polytechnique.edu
[‡]friz@math.tu-berlin.de
[§]paolo.pigato@uniroma2.it

1 Introduction

In rough stochastic volatility models, volatility is driven by a fractional noise, in the rough regime of Hurst parameter H less than $1/2$. With no claim of completeness, we mention econometric evidence [37, 33, 11], market microstructure foundations, [20], efficient numerical methods, and simulations schemes [7, 10, 48, 21], including deep learning algorithms [9, 39].

This work is concerned with pricing under rough volatility, a key feature of which, well-adapted to the steep volatility skews seen in equity option markets, is the power-law behavior of the short-dated *implied volatility* at-the-money (ATM) skew:

$$\mathcal{S}_{\mathrm{BS}} \sim (\mathrm{const}) t^{H-1/2} \, ;$$

references include [1, 30, 4, 6, 22, 19, 31, 25, 41, 32, 8]. More specifically, we consider here - to the best of our knowledge for the first time - the Dupire *local volatility* [18, 17] generated by rough volatility models, using a Gyongy type projection [43, 14] and study its short-time behavior in a large deviations regime. In the context of *implied volatility*, such a regime was pioneered in [22], with instantaneous stochastic volatility process given as $v(t, \omega) = \sigma(W_t^H(\omega))$, i.e., as some explicit function of a fractional Brownian motion (fBm). In particular, no rough or Volterra stochastic differential equations need to be solved. We shall work under the same structural assumption as [22], although under less restrictive growth assumptions, such as to include the popular rough Bergomi model [4] where σ has exponential form. Postponing detailed recalls and precise definitions, our main result (Theorem 3.3) states that,

$$\sigma_{\mathrm{loc}}\big(t, y\, t^{1/2-H}\big) \to \sigma\big(\widehat{h}_1^y\big) \quad \text{as } t \downarrow 0,$$

where $\widehat{h}^y$ is related to a minimization problem, similar to a geodesic in Riemannian geometry. Our analytic understanding is sufficiently fine to exploit it on the one hand for numerical tests (discussed in Section 4) and on the other hand to derive further analytic results (formulated in Sections 2 and 3, with proofs left to Section 5) including the blowup, when $H < 1/2$, of the *local volatility skew* in the short-dated limit,

$$\mathcal{S}_{\mathrm{loc}} \sim (\mathrm{const}) t^{H-1/2},$$

see Corollary 3.4 below for a precise statement and information on the constant. This finding is consistent with [31] where it is shown, amongst others, that in "regular" local-stochastic vol models, which amounts to a regularity assumption on σ_{loc}, the implied volatility skew does not explode. The regularity of σ_{loc} is violated here in the sense that $\mathcal{S}_{\mathrm{loc}}$ is infinite at $t = 0$. This is also consistent with [50, 29] where it is shown that a "singular" σ_{loc} can indeed produce exploding implied skews.

A further interesting consequence, also part of Corollary 3.4, is then that the $1/2$-rule of thumb from practitioners, shown to hold in classical stochastic volatility models [12, 35], actually *fails* and is replaced, again in the short-dated limit, by what we may call the $1/(H + 3/2)$-rule,

$$\frac{\mathcal{S}_{\mathrm{BS}}}{\mathcal{S}_{\mathrm{loc}}} \to \frac{1}{H + 3/2}.$$

As a sanity check, for Hurst parameter $H = 1/2$ we are in a diffusive regime and then indeed fall back to the $1/2$-rule.

Techniques and further discussion. Our analysis is based on a mixture of *large deviations*, e.g., see [26] for a recent collection with many references, *Malliavin calculus* [3, 49, 23], and last not least ideas from *rough paths* and *regularity structures* techniques, following [5, 24, 25]; see also section 14.6 in [28]. In order to deal with the case $H < 1/2$, we cannot rely on previously used methods in diffusion settings such as [53, 16]. Local volatility in classical stochastic volatility models, including Heston, is discussed in many books on volatility modeling, [35] remains a key reference. Rigorous asymptotic results include [36, 15, 16]. In affine forward variance models, including rough Heston [38], it is conceivable that saddle-point-based techniques, in the spirit of [15] can be employed to study local volatility asymptotics. The bottleneck in such an approach seems however the lack of explicit knowledge of the moment generating function, only given implicitly via convolution Riccati equations.

2 The modeling framework

We assume $S_0 = 1$ and that the log price $X_t := \log S_t$ satisfies

$$\begin{aligned}
\mathrm{d}X_t &= -\frac{1}{2}V_t\mathrm{d}t + \sqrt{V_t}\left(\rho\,\mathrm{d}W_t + \overline{\rho}\,\mathrm{d}\overline{W}_t\right) \\
V_t &= \sigma^2(\widehat{W}_t)
\end{aligned} \tag{2.1}$$

with 'volatility function' $\sigma : \mathbb{R} \to \mathbb{R}$. We shall assume σ to be smooth, subject to mild growth conditions given below, such as to cover rough Bergomi type situations where $\sigma(x) = \sigma_0 \exp(\eta x)$.

Assumption 2.1. There exist $c_1, c_2, c_3, c_4 > 0$ such that for all $x \in \mathbb{R}$,

$$c_1 e^{-c_2|x|} \leq \sigma(x), \tag{2.2}$$

$$\sigma(x) \leq c_3 e^{c_4|x|}. \tag{2.3}$$

We take $\rho^2 + \overline{\rho}^2 = 1$, with $\rho \in (-1, 1)$. We denote $\mathrm{W} = (W, \overline{W})$ where $W, \overline{W}$ are two independent standard Brownian motions. These are used to construct

$$\widetilde{W}_t = \overline{\rho}\,\overline{W}_t + \rho W_t \quad \text{and} \quad \widehat{W}_t = (K * \dot{W})_t = \int_0^t K(t,s)\mathrm{d}W_s, \tag{2.4}$$

with

$$K(t,s) = \sqrt{2H}(t-s)^{H-1/2} \qquad \text{for } t > s$$

and $K(t,s) = 0$ otherwise, so that $\widetilde{W}$ is again a standard Brownian motion (ρ-correlated with W) whereas $\widehat{W}$ is a Riemann–Liouville fBm with Hurst index $H \leq 1/2$, i.e., the self-similar Gaussian Volterra process in (2.4).

We will use analogous notations for Cameron–Martin paths $\mathrm{h} = (h, \overline{h})$, so that $\widetilde{h} = \overline{\rho}\,\overline{h} + \rho\,h$, and $\widehat{h}_t = (K^H * \dot{h})_t = \int_0^t K(t,s)\mathrm{d}h_s$. We denote H^1 the Cameron–Martin space and $\|\cdot\|_{H^1}$ the Cameron–Martin norm $\|\mathrm{h}\|_{H^1}^2 = \int_0^1 (\dot{h}^2 + \dot{\overline{h}}^2)\mathrm{d}t$.

3 Mathematical setting and results

The time-scaling property of the Gaussian process $(W, \overline{W}, \widehat{W})$ underlying the model (2.1) yields $X_{\varepsilon^2} \overset{\text{law}}{=} X_1^\varepsilon$ for every $\varepsilon > 0$, where X_1^ε satisfies

$$X_1^\varepsilon = \int_0^1 \sigma\left(\varepsilon^{2H}\widehat{W}_s\right)\varepsilon\,\mathrm{d}\left(\overline{\rho}\,\overline{W} + \rho\,W\right)_s - \frac{1}{2}\varepsilon^2 \int_0^1 \sigma^2\left(\varepsilon^{2H}\widehat{W}_s\right)\mathrm{d}s. \tag{3.1}$$

Forde and Zhang proved in [22], albeit under different technical conditions on the volatility function, that a small noise Large Deviation Principle (LDP) holds for the family $\varepsilon^{2H-1}X_1^\varepsilon$ (hence for $\varepsilon^{2H-1}X_{\varepsilon^2}$) as $\varepsilon \to 0$, with speed ε^{4H} and rate function

$$\Lambda(y) := \inf_{\mathrm{h} \in H^1}\left\{\frac{1}{2}\int_0^1 \dot{h}^2\mathrm{d}t + \frac{1}{2}\int_0^1 \dot{\overline{h}}^2\,\mathrm{d}t : \Phi_1(\mathrm{h}) = y\right\} = \frac{1}{2}\|\mathrm{h}^y\|_{H^1}^2\,, \tag{3.2}$$

where $\Phi_1(\cdot)$ is the first component of the Itô map of (2.1), such that $X_t = \Phi_t(\mathrm{W})$, and h^y is a minimizer of the control problem defining $\Lambda(y)$. From the LDP (3.2), we have

$$-\varepsilon^{4H}\log\mathbb{P}\left(X_1^\varepsilon \geq y\varepsilon^{1-2H}\right) \to \Lambda(y) = \frac{1}{2}\|\mathrm{h}^y\|_{H^1}^2, \text{ for } y \geq 0 \text{ as } \varepsilon \downarrow 0, \tag{3.3}$$

$$-\varepsilon^{4H}\log\mathbb{P}\left(X_1^\varepsilon \leq y\varepsilon^{1-2H}\right) \to \Lambda(y) = \frac{1}{2}\|\mathrm{h}^y\|_{H^1}^2, \text{ for } y \leq 0, \text{ as } \varepsilon \downarrow 0, \tag{3.4}$$

and this small-noise LDP eventually translates to a short-time LDP for the process X_{ε^2}. This result was proved in the case where $V_t = \sigma(\widehat{W}_t)$ in [22], and then extended to the possible time dependence of the form $V_t = \sigma(\widehat{W}_t, t^{2H})$ in [24, section 7.3] (see also Remark 3.7).

The short-time result for call and put prices reads as follows (see [22, Corollary 4.13])

$$-t^{2H}\log\mathbb{E}\left[(e^{X_t} - e^{y\,t^{1/2-H}})^+\right] \to \Lambda(y) = \frac{1}{2}\|\mathrm{h}^y\|_{H^1}^2, \text{ for } y > 0 \text{ as } t \downarrow 0, \tag{3.5}$$

$$-t^{2H}\log\mathbb{E}\left[(e^{y\,t^{1/2-H}} - e^{X_t})^+\right] \to \Lambda(y) = \frac{1}{2}\|\mathrm{h}^y\|_{H^1}^2, \text{ for } y < 0 \text{ as } t \downarrow 0, \tag{3.6}$$

where h^y is as above. Let us also recall that these option price asymptotics imply the following asymptotic formula for the Black–Scholes implied volatility (notation: σ_{BS}), which can be seen as a "rough" version of the Berestycki–Buscat–Florent (BBF) formula [13]:

$$\sigma_{\mathrm{BS}}^2(t, yt^{1/2-H}) \to \chi^2(y) := \frac{y^2}{2\Lambda(y)} \quad \text{for } y \neq 0 \text{ as } t \downarrow 0. \tag{3.7}$$

Remark 3.1 (Precise conditions for the LDP, call price asymptotics, and implied volatility asymptotics). The exponential growth condition (2.3) is no obstruction for an LDP to hold for the model (2.1), as was shown in [5, 40], weakening the linear growth condition first required in [22]. Moreover, while the put price asymptotics (3.6) always holds, the unboundedness of the call option payoff requires some additional condition for (3.5) to hold: with reference to [24, Assumption A2], we will assume the following "1+ moment condition" whenever necessary:

Assumption 3.2. There exists $p > 1$ such that $\limsup_{\varepsilon \to 0} \mathbb{E}[e^{pX_1^\varepsilon}] < \infty$.

Following [24, Lemma 4.7], Assumption 3.2 is true under the following stronger, but more explicit, condition: the process $S_t = e^{X_t}$ is a martingale, and there exist $p > 1$ and $t > 0$ such that $\mathbb{E}[S_t^p] < \infty$. It is known that such a condition on the moments of e^{X_t} is satisfied when σ has linear growth, cf. [22], while in the case $H = 1/2$, the same is true under much weaker assumptions (σ of exponential growth and $\rho < 0$ is enough, see [52, 47]). We expect similar results to hold for $H < 1/2$, but they have not been proved yet; see the partial results available in [34, 41].

The Markovian projection of the instantaneous variance V_t (see [43],[14, Corollary 3.7]) within the model (2.1) is defined by

$$\sigma_{\mathrm{loc}}^2(t, k) := \mathbb{E}\big[V_t | X_t = k\big] \quad \text{for every } t > 0 \text{ and } k \in \mathbb{R}. \tag{3.8}$$

It follows from references [43, 14] that the dynamics of the resulting local volatility model are weakly well-posed; see also [27] for a generic regularization scheme obtained by time-shifting the local volatility surface (a procedure that we do not require here).

We now present our main result. We prove that the local volatility function (3.8) satisfies the following short-time asymptotics.

Theorem 3.3 (Markovian projection at the LDP regime). *Let Assumption 2.1 be in force. Then, the Markovian projection in the model (2.1) satisfies, for every $y \in \mathbb{R} \setminus \{0\}$ small enough,*

$$\sigma_{\mathrm{loc}}^2\big(t, y\, t^{1/2-H}\big) = \mathbb{E}\big[V_t | X_t = yt^{1/2-H}\big] \to \sigma^2\big(\widehat{h}_1^y\big) \quad \text{as } t \downarrow 0, \tag{3.9}$$

where we recall that $\widehat{h}_t^y = \int_0^t K(t,s)\mathrm{d}h_s^y$ and $\mathrm{h}^y = (h^y, \overline{h}^y)$ is the minimizer of the rate function in (3.2).

Let us stress that the asymptotics (3.9) for the local volatility function holds under the mild growth conditions of Assumption 2.1, while we do not require the 1+ moment condition of Assumption 3.2.

3.1 Local volatility skew and the new $\frac{1}{H+3/2}$ rule

Let us write $\sim$ for asymptotic equivalence as $t \to 0$. Let us denote

$$\Sigma(y) := \sigma\left(\widehat{h}_1^y\right)$$

the limiting function in (3.9), and consider the following finite-difference approximations of the local and implied volatility skew

$$\mathcal{S}_{\mathrm{loc}}(t,y) := \frac{\sigma_{\mathrm{loc}}(t, y\, t^{1/2-H}) - \sigma_{\mathrm{loc}}(t, -yt^{1/2-H})}{2y\, t^{1/2-H}}, \tag{3.10}$$

$$\mathcal{S}_{\mathrm{BS}}(t,y) := \frac{\sigma_{\mathrm{BS}}(t, y\, t^{1/2-H}) - \sigma_{\mathrm{BS}}(t, -yt^{1/2-H})}{2y\, t^{1/2-H}}. \tag{3.11}$$

Then, we have the following

Corollary 3.4 (Local vol skew and the new $\frac{1}{H+3/2}$ rule). *Let $\rho \neq 0$. Let Assumption 2.1 be in force. Then, for $y \in \mathbb{R} \setminus \{0\}$ small enough,*

$$\mathcal{S}_{\mathrm{loc}}(t,y) \sim \frac{\Sigma(y) - \Sigma(-y)}{2y} \frac{1}{t^{1/2-H}} \tag{3.12}$$

as $t \to 0$. Under the additional moment condition in Assumption 3.2,

$$\lim_{y \to 0} \lim_{t \to 0} \frac{\mathcal{S}_{\mathrm{BS}}(t,y)}{\mathcal{S}_{\mathrm{loc}}(t,y)} = \frac{1}{H + 3/2}. \tag{3.13}$$

In the case $\rho = 0$, we have $\mathcal{S}_{\mathrm{BS}}(t,y) = 0$ and $\mathcal{S}_{\mathrm{loc}}(t,y) = 0$ for every t.

In our numerical experiments in section 4, we estimate the exact ATM local volatility skew $\frac{1}{2}\partial_k \sigma_{\mathrm{loc}}(t,k)\big|_{k=0}$ in the rough Bergomi model (4.1), and find perfect agreement with Corollary 3.4. The model local volatility skew can be observed in Figure 1, and the ratio of the implied volatility skew over the local volatility skew in Figure 2.

Remark 3.5. When $H = 1/2$, we are back to the classical $1/2$ skew rule, see [12, 36].

Remark 3.6 (The short-time harmonic mean formula). When expressed in terms of an implied volatility σ_{BS}, Dupire's formula for local volatility reads

$$\sigma_{\mathrm{loc}}(t,k)^2 = \frac{\sigma_{\mathrm{BS}}(t,k) + 2\, t\, \partial_t \sigma_{\mathrm{BS}}(t,k)}{\left(t\, \partial_{kk}\sigma_{\mathrm{BS}} - \frac{1}{4}t^2\, \sigma_{\mathrm{BS}}(\partial_k \sigma_{\mathrm{BS}})^2 + \frac{1}{\sigma_{\mathrm{BS}}}\left(1 - \frac{k\, \partial_k \sigma_{\mathrm{BS}}}{\sigma_{\mathrm{BS}}}\right)^2\right)(t,k)} \tag{3.14}$$

provided that σ_{BS} is sufficiently smooth for all the partial derivatives to make sense. Formally taking $t \to 0$ inside (3.14) and assuming that the partial derivatives $\partial_t \sigma_{\mathrm{BS}}$, $\partial_k \sigma_{\mathrm{BS}}$ and $\partial_{kk}\sigma_{\mathrm{BS}}$ remain bounded, one obtains

$$\sigma_{\mathrm{loc}}(0,k)^2 = \frac{\sigma_{\mathrm{BS}}(0,k)^2}{\left(1 - k\frac{\sigma'_{\mathrm{BS}}(0,k)}{\sigma_{\mathrm{BS}}(0,k)}\right)^2}. \tag{3.15}$$

The ordinary differential equation (3.15) can be used to reconstruct the function $\sigma_{\mathrm{BS}}(0, \cdot)$ from $\sigma_{\mathrm{loc}}(0, \cdot)$ and it is solved by the harmonic mean function

$$H(t, k) = \frac{1}{\frac{1}{k} \int_0^k \frac{\mathrm{d}y}{\sigma_{\mathrm{loc}}(t, y)}}, \tag{3.16}$$

evaluated at $t = 0$. The rigorous counterpart of this formal argument, that is the asymptotic equivalence $\sigma_{\mathrm{BS}}(t, k) \sim H(t, k)$, known as the "harmonic mean formula" or BBF formula was proven in [12] under the assumption that the local volatility surface σ_{loc} is bounded and uniformly continuous in a neighborhood of $t = 0$. It is straightforward to see that the harmonic mean satisfies the property $\partial_k H(t, k)\big|_{k=0} = \frac{1}{2} \partial_k \sigma_{\mathrm{loc}}(t, k)\big|_{k=0}$. Therefore, if we assume that the short-time approximation property $\sigma_{\mathrm{BS}}(t, k) \approx H(t, k)$ also holds for the first derivatives with respect to k, as a consequence we obtain the 1/2 short-time skew rule $\partial_k \sigma_{\mathrm{BS}}(t, 0) \sim \frac{1}{2} \partial_k \sigma_{\mathrm{loc}}(t, 0)$ that we referred to in Remark 3.5.

Corollary 3.4 entails that the formal argument above does not hold anymore for the implied and local volatility surfaces generated by a rough stochastic volatility model. Notably, the boundedness of the partial derivatives $\partial_t \sigma_{\mathrm{BS}}$, $\partial_k \sigma_{\mathrm{BS}}$ and $\partial_{kk} \sigma_{\mathrm{BS}}$, and the uniform continuity of the local volatility surface, fall short – but in such a way that the limit of the skew ratio $\frac{\mathcal{S}_{\mathrm{BS}}}{\mathcal{S}_{\mathrm{loc}}}$ can still be identified and explicitly computed (see related numerical tests in Figure 4).

Remark 3.7 (Time-dependent volatility function). The rough Bergomi model [4] comes with instantaneous variance

$$\xi(t) \exp\left(\eta x - \frac{\eta^2}{2} t^{2H}\right) \sim \xi(0) \exp(\eta x) =: \sigma^2(x)$$

as $t \downarrow 0$. We could have proved Theorem 3.3 and Corollary 3.4 in greater generality, with $V_t = \sigma^2(\widehat{W}_t, t)$, provided the dependence with respect to t in $\sigma = \sigma(x, t)$ is sufficiently smooth such as not to affect the local analysis that underlies the proof. This is more subtle in the case of rough Bergomi where t^{2H} fails to be smooth at $t = 0^+$ when $H < 1/2$. Even so, we discussed in [24] how to adjust the arguments to obtain exact asymptotics, the same logic applies here.

4 Numerical tests

We wish to estimate the conditional expectation (3.8) for some specific instance of the model (2.1), using Monte Carlo simulation. We consider the rough Bergomi model [4], for which the instantaneous variance process is given by

$$V_t = \xi_0 \exp\left(\eta \widehat{W}_t - \frac{\eta^2}{2} t^{2H}\right) = \xi_0 \exp\left(\eta \int_0^t \sqrt{2H}(t - s)^{H-1/2} \mathrm{d}W_s - \frac{\eta^2}{2} t^{2H}\right), \tag{4.1}$$

where $\xi_0 = V_0$ is the spot variance and η a parameter that tunes the volatility of variance. Note that, strictly speaking, Theorem 3.3 and Corollary 3.4 do not apply to the model above, because of the time dependence in the volatility function $\sigma(x, t) = \sqrt{\xi_0} \exp\left(\frac{\eta}{2} x - \frac{\eta^2}{4} t^{2H}\right)$. In light of the discussion in Remark 3.7, we can expect our asymptotic results to hold for such a time-dependent volatility function as well, which is in line with the output of our numerical experiments below.

For a given time horizon $T > 0$ and a number $N \in \mathbb{N}^*$ of time-steps, the random vector $(\log V_{t_k})_{1 \le k \le N}$, $t_k = k\frac{T}{N}$, has a multivariate Gaussian distribution with known mean and variance, see for example [4], and can therefore be simulated exactly. We use the standard simulation method

for Gaussian vectors based on a Cholesky factorization of the covariance matrix. Of course, this method has a considerable complexity – cost $\mathcal{O}(N^3)$ for the Cholesky factorization and $\mathcal{O}(N^2)$ for the matrix multiplication required to get one sample of $(V_{t_k})_{0 \le k \le N}$ – but our focus is on the accuracy of our estimations, rather than on their computational time. We construct approximate samples of the log-asset price $X_T = -\frac{1}{2} \int_0^T V_t dt + \int_0^T \sqrt{V_t}(\rho dW_t + \overline{\rho} d\overline{W}_t)$ using a forward Euler scheme on the same time-grid

$$X_T^N = -\frac{T}{2N} \sum_{k=0}^{N-1} V_{t_k} + \sum_{k=0}^{N-1} \sqrt{V_{t_k}} \Big(\rho(W_{t_{k+1}} - W_{t_k}) + \overline{\rho}(\overline{W}_{t_{k+1}} - \overline{W}_{t_k}) \Big).$$

Therefore, we obtain M i.i.d. approximate Monte Carlo samples $(X_T^{N,m}, V_T^m)_{1 \le m \le M}$ of the couple (X_T^N, V_T), from which our estimators of the implied volatility and local volatility (3.8) are constructed, as detailed below. Since our goal is to check the asymptotic statements appearing in Theorem 3.3 and Corollary 3.4, we will consider a large number N of discretization steps and a large number M of Monte Carlo samples in order to increase the precision of the estimates we use as a benchmark. We estimate out-of-the-money put and call option prices by standard empirical means and evaluate the corresponding implied volatilities $\sigma_{\mathrm{BS}}(T, K)$ by Newton's search.

The rough Bergomi model (4.1) parameters we used in our experiments are $S_0 = 1, \eta = 1.0, \rho = -0.7$, and $\xi_0 = 0.235^2$. We tested three different values of $H \in (0, 1/2]$, namely $H = 0.1, 0.3$, and 0.5. We used $M = 1.5 \times 10^6$ Monte Carlo samples and $N = 500$ discretization points.

4.1 Local and implied volatility estimators

In this section, we present in detail the estimators we have implemented for the target objects: the at-the-money implied volatility skew $\partial_k \sigma_{\mathrm{BS}}(t, k)|_{k=0}$, the local volatility function (or Markovian projection) $\sigma_{\mathrm{loc}}(\cdot, \cdot)$ in (3.8), and the local volatility skew $\partial_k \sigma_{\mathrm{loc}}(t, k)|_{k=0}$.

Estimator of the implied volatility skew. A representation of the first derivative $\partial_k \sigma_{\mathrm{BS}}(t, k)$ can be obtained by differentiating the equation defining the implied volatility σ_{BS} with respect to the log-moneyness k. More precisely, denoting $C_{\mathrm{BS}}(k, v)$ the Black–Scholes price of a call option with log-moneyness k and total volatility parameter $v = \sqrt{t} \sigma$, we have

$$\mathbb{E}\big[(S_0 e^{X_t} - S_0 e^k)^+\big] = C_{\mathrm{BS}}\big(k, \sqrt{t} \sigma_{\mathrm{BS}}(t, k)\big), \tag{4.2}$$

for all k and t. Taking the derivative at both sides of (4.2) with respect to k and using the expressions of the first-order Black–Scholes greeks $\partial_k C_{\mathrm{BS}}(k, v)$ and $\partial_v C_{\mathrm{BS}}(k, v)$, we have

$$\partial_k \sigma_{\mathrm{BS}}(t, k) = \frac{-\partial_k C_{\mathrm{BS}}(k, v) - S_0 e^k \, \mathbb{P}\big(X_t \ge k\big)}{\sqrt{t} \, \partial_v C_{\mathrm{BS}}(k, v)} \bigg|_{v = \sqrt{t} \sigma_{\mathrm{BS}}(t, k)}$$
$$= \frac{N\big(d_2(k, v)\big) - \mathbb{P}\big(X_t \ge k\big)}{\sqrt{t} \, \phi\big(d_2(k, v)\big)} \bigg|_{v = \sqrt{t} \sigma_{\mathrm{BS}}(t, k)},$$

where $d_2(k, v) = -\frac{k}{v} - \frac{v}{2}$, and ϕ (resp. N) denotes the standard Gaussian density (resp. cumulative distribution). The representation above for the implied volatility skew allows to avoid finite difference methods and only requires estimating $\sigma_{\mathrm{BS}}(t, k)$ and $\mathbb{P}(X_t \ge k)$, which we can do with the same Monte Carlo sample, in order to estimate $\partial_k \sigma_{\mathrm{BS}}(t, k)$ (and therefore, in particular, the at-the-money skew $\partial_k \sigma_{\mathrm{BS}}(t, 0)$).

8

Estimator of the local volatility function. Given the Monte Carlo samples $(X_T^{N,m}, V_T^m)_{1 \le m \le M}$ of the couple (X_T^N, V_T), the conditional expectation (3.8) defining the local volatility function can be estimated by appealing to several different regression methods, see, e.g., [54, 45]. We have implemented and benchmarked two different estimators: on the one side, a kernel regressor, already applied to evaluate the Markovian projection within the celebrated particular calibration algorithm [42], and on the other side, an alternative estimator based on the explicit knowledge of the conditional law of $(X_t, V_t)|(W_s)_{s \le t}$.

Our kernel regressor is the Nadaraya–Watson estimator with bandwidth δ,

$$\sigma_{\text{loc}}^2(t, k) = \mathbb{E}\left[V_t | X_t = k\right] \approx \frac{\sum_{m=1}^M V_t^m K_\delta\left(X_t^{N,m} - k\right)}{\sum_{m=1}^M K_\delta\left(X_t^{N,m} - k\right)}. \tag{4.3}$$

We used a Gaussian kernel $K_\delta(x) = \exp(-\delta x^2)$ in our tests.

On the other hand, it is a standard fact that conditionally on $\mathcal{F}_t = \sigma(W_u : u \le t)$, the instantaneous variance V_t is known, while the log-price X_t is normally distributed with mean $-\frac{1}{2}\int_0^t V_s \mathrm{d}s + \rho \int_0^t \sqrt{V_s}\mathrm{d}W_s$ and variance $(1 - \rho^2)\int_0^t V_s \mathrm{d}s$, see, e.g., [51]. This property yields a representation of the Markovian projection $\sigma_{\text{loc}}(\cdot, \cdot)$ as the ratio of two expectations (which, incidentally, has already been exploited in [44] in the framework of a calibration strategy of local stochastic volatility models – prior to the particular algorithm [42]):

$$\sigma_{\text{loc}}^2(t, k) = \mathbb{E}\left[V_t | X_t = k\right] = \frac{\mathbb{E}\left[V_t \, \Pi_t(k)\right]}{\mathbb{E}\left[\Pi_t(k)\right]} \tag{4.4}$$

where

$$\Pi_t(k) = \frac{1}{\sqrt{\int_0^t V_s \mathrm{d}s}} \exp\left(-\frac{1}{2(1 - \rho^2)\int_0^t V_s \mathrm{d}s}\left(k + \frac{1}{2}\int_0^t V_s \mathrm{d}s - \rho \int_0^t \sqrt{V_s}\mathrm{d}W_s\right)^2\right).$$

Estimator of the local volatility skew. Differentiating the right-hand side of (4.4) with respect to k, we obtain a representation of $\partial_k \sigma_{\text{loc}}(t, k)$:

$$\partial_k \sigma_{\text{loc}}(t, k) = \frac{\frac{\partial}{\partial k}\left(\frac{\mathbb{E}[V_t \Pi_t]}{\mathbb{E}[\Pi_t]}\right)}{2\,\sigma_{\text{loc}}(t, k)} = \frac{\mathbb{E}\left[V_t \Pi_t\right] \mathbb{E}\left[\frac{U}{\int_0^t V_s \mathrm{d}s}\Pi_t\right] - \mathbb{E}\left[\frac{U}{\int_0^t V_s \mathrm{d}s}\Pi_t V_t\right] \mathbb{E}\left[\Pi_t\right]}{2(1 - \rho^2)\mathbb{E}\left[V_t \Pi_t\right]^{1/2} \mathbb{E}\left[\Pi_t\right]^{3/2}}, \tag{4.5}$$

where Π_t is a shorthand for $\Pi_t(k)$, $U = U(k) = k + \frac{1}{2}\int_0^t V_s \mathrm{d}s - \rho \int_0^t \sqrt{V_s}\mathrm{d}W_s$, and $\frac{\partial \Pi_t}{\partial k} = -\frac{U}{(1 - \rho^2)\int_0^t V_s \mathrm{d}s}\Pi_t$. All the expectations appearing in (4.4) and (4.5) can be estimated based on the exact simulation of the discretized variance path $(V_{t_k})_{1 \le k \le N}$; we approximate the integrals $\int_0^t V_s \mathrm{d}s$ and $\int_0^t \sqrt{V_s}\mathrm{d}W_s$ using left-point Euler schemes. Note that the resulting non-parametric estimators based on the representations (4.4) and (4.5) do not contain any kernel bandwidth or other hyper-parameters to be tuned. This is a clear advantage with respect to (4.3). We have nevertheless tested both estimators (4.3) and (4.4) for the Markovian projection function, and found perfect agreement between the two in our tests – in other words, the local volatilities and local volatility skews computed with the two different methods would be indistinguishable in Figures 1 and 3.

In Figure 1, we plot the term structure of the ATM implied and local volatility skews, for three different values of H and maturities up to $T = 0.5$ years. As pointed out in the Introduction and

in Section 2, the power-law behavior of the ATM implied volatility skew generated by the rough Bergomi model is already well-known; on the other hand, the power-law behavior observed for the local volatility skew in Figure 1 is (to the best of our knowledge) new, and consistent with Corollary 3.4. Figure 2 shows the ratio of the implied volatility ATM skew over the local volatility ATM skew, that is the ratio of the curves observed in Figure 1, for the different values of H: the numerical results are in very good agreement with the "$\frac{1}{H+3/2}$ rule" announced in Corollary 3.4. Additionally, we note that the ratio of the two skews seems to be rather stable – its value is almost constant for maturities up to $T = 0.5$ years, with our parameter setup.

4.2 Short-dated local volatility

Theorem 3.3 gives the asymptotic behavior of $\sigma_{\mathrm{loc}}\big(T, y\, T^{1/2-H}\big)$ as T becomes small. Since y is allowed to vary around the at-the-money point $y = 0$, we can check whether the limit (3.9) holds for the function $y \mapsto \widehat{\sigma}_{\mathrm{loc}}(T, y) := \sigma_{\mathrm{loc}}\big(T, y\, T^{1/2-H}\big)$, that is the whole local volatility smile rescaled with maturity. The computation of the limiting function $\sigma\big(\widehat{h}_1^y\big)$ requires evaluating the Cameron-Martin path $h^y \in H^1$ that minimizes the rate function in (3.2) for a given y. We follow the procedure already exploited in [22] and [25, section 5.1] : it can be shown, see [22], that the rate function satisfies the alternative representation $\Lambda(y) = \inf\Big\{ \frac{(y-\rho\,G(h))^2}{2\,\overline{\rho}^2 F(h)} + \frac{1}{2}\langle\dot{h},\dot{h}\rangle\,:\,\dot{h}\,\in\,L^2(0,1)\Big\}$, with $F(h) = \langle\sigma^2(\widehat{h}),1\rangle = \int_0^1 \sigma^2(\widehat{h}_t)\mathrm{d}t$ and $G(h) = \langle\sigma(\widehat{h}),\dot{h}\rangle = \int_0^1 \sigma(\widehat{h}_t)\dot{h}_t\mathrm{d}t$. This alternative representation yields the rate function under the form of an unconstrained optimization problem (as opposed to the constrained optimization in (3.2)), which can then be approximately solved by the projection of a one-dimensional path h over an orthonormal basis $\{\dot{e}_n\}_{n\geq 1}$ of L^2, $\dot{h}_t = \sum_{n\geq 1} a_n \dot{e}_n(t)$. In practice, we truncate the sum at a certain order N and minimize over the coefficients $(a_n)_{1\leq n\leq N}$; we obtain an approximation of the minimizer h^y and therefore of $\widehat{h}_t^y = (K^H * \dot{h}^y)_t = \int_0^t K(t,s)\dot{h}_s\mathrm{d}s$. We chose the Fourier basis $\big\{\dot{e}_1(t) = 1,\ \dot{e}_{2n}(t) = \sqrt{2}\cos(2\pi n\,t),\ \dot{e}_{2n+1}(t) = \sqrt{2}\sin(2\pi n\,t), n \in \mathbb{N}\backslash\{0\}\big\}$ in our experiments, and observed that truncation of the sum at $N = 8$ provides a good accuracy. The results for the rough Bergomi model are displayed in Figure 3, where the function $\widehat{\sigma}_{\mathrm{loc}}(T,y)$ is indeed seen to approach its limit $\sigma\big(\widehat{h}_1^y\big)$ when maturity decreases from $T = 0.5$ to $T = 0.05$. The residual error term $\widehat{\sigma}_{\mathrm{loc}}(T,y) - \sigma\big(\widehat{h}_1^y\big)$ is seen to depend on H, with lower values of H being associated with higher errors.

Extrapolation of local volatility surfaces. Eventually, Theorem 3.3 provides us with an extrapolation recipe for local volatilities for very short maturities: fixing a (small) maturity T and a log-moneyness level k, formally plugging $y = \frac{k}{T^{1/2-H}}$ in (3.9) we obtain

$$\sigma_{\mathrm{loc}}\big(T,k\big) \approx \sigma\Big(\widehat{h}_1^y\Big)\Big|_{y=\frac{k}{T^{1/2-H}}}\,.$$

The limiting function $\sigma\big(\widehat{h}_1^y\big)\big|_{y=\frac{k}{T^{1/2-H}}}$ can therefore be used to extrapolate a local volatility surface that is consistent with the behavior implied by a rough volatility model.

As an application, consider the calibration of a Local-Stochastic Volatility model (LSV) to an option price surface, for example using the particle method of Guyon and Henry-Labordère[42]. The LSV model can be obtained by the decoration of a naked rough volatility model, which amounts to

enhance the rough vol model (2.1) for $S_t = S_0 e^{X_t}$ with a leverage function $l(t, S)$,

$$dS_t = S_t \, l(t, S_t) \sqrt{V_t} \left(\rho \, dW_t + \sqrt{1 - \rho^2} \, d\overline{W}_t \right).$$

Given the spot variance process V, the LSV model calibrated to a given Dupire local volatility surface σ_{Dup} (see [42] for details) corresponds to

$$l(t, S_t) = \frac{\sigma_{\mathrm{Dup}}(t, S_t)}{\sqrt{\mathbb{E}[V_t | S_t]}}.$$

In general, one wishes the leverage function $l(t, S)$ to be a small correction to the original stochastic volatility model (in other words: as close as possible to $l \equiv 1$). In practice, the local volatility σ_{Dup} coming from market data has to be extrapolated for values of t smaller than the shorter observed maturity, and the choice of the extrapolation method is up to the user. If for small t, the chosen extrapolation $\sigma_{\mathrm{Dup}}(t, K)$ is qualitatively too different from the behavior of the conditional expectation $\mathbb{E}[V_t | S_t = K]$ in the rough volatility setting (for example, more specifically: the ATM skew of σ_{Dup} is far from the power-law (3.12)), then the leverage function will have to compensate, hence deviating from the unit function. Under the pure rough volatility model ($l \equiv 1$), Theorem 3.3 and Corollary 3.4 describe the behavior of the markovian projection $\mathbb{E}[V_t | S_t]$ for small t: eventually, these statements give hints on how $\sigma_{\mathrm{Dup}}(t, \cdot)$ should be extrapolated in order for $l(t, \cdot)$ not to deviate too much from the unit function.[1]

Failure of the harmonic mean asymptotic formula under rough volatility. In Remark 3.6, we pointed out that, as a consequence of the general $\frac{1}{H+3/2}$ skew rule (as opposed to the $1/2$ rule) in Corollary 3.4, the harmonic mean asymptotic formula $\sigma_{\mathrm{BS}}(T, k) \sim H(T, k)$ as $T \to 0$, see (3.16), is expected not to hold for $H \neq 1/2$ (without any contradiction with the statements in [12], which require regularity conditions on local volatility surface that are not satisfied in the rough volatility setting, see our discussion in Remark 3.6). In other words, we do not expect the harmonic mean of the local volatility $H(T, k) = \frac{1}{\frac{1}{k} \int_0^k \frac{dy}{\sigma_{\mathrm{loc}}(T, y)}}$ to be a good approximation of the implied volatility when maturities become small when the involved volatility surfaces are generated by a rough vol model. Having constructed estimators (4.3) and (4.4) of the local volatility function under the rough Bergomi model, we are also able to approximate (with an additional deterministic quadrature) the harmonic mean $H(T, k)$, and to compare the output with the implied volatility smile. The results are shown in Figure 4, for three different values of H. As expected, when $H = 0.5$ we observe (upper left panel) that the implied volatility $\sigma_{\mathrm{BS}}(T, k)$ approaches the harmonic mean $H(T, k)$ when maturity decreases from $T = 0.45$ to $T = 0.05$, and the at-the-money slopes are also seen to agree. The convergence is even more apparent in the upper right figure, where the ratio $\frac{\sigma_{\mathrm{BS}}(T,k)}{H(T,k)}$ is seen to monotonically converge to one. This behavior should be compared with the one in the two bottom figures, where the rough case $H = 0.1$ is considered (the case $H = 0.3$ being intermediate between the other two): now, when maturity decreases, the implied volatility smile does not seem to approach the harmonic mean $H(T, k)$ anymore, and in particular, the slopes of the two curves are seen to considerably deviate from each other. This phenomenon is even more clear in the bottom

---

[1] We thank Andrea Pallavicini and Roberto Longoni from Banca Intesa Sanpaolo for interesting and stimulating discussions on this topic.

right figure, where the ratio $\frac{\sigma_{\mathrm{BS}}(T,k)}{H(T,k)}$ has a completely different behavior with respect to the diffusive case $H = 0.5$.
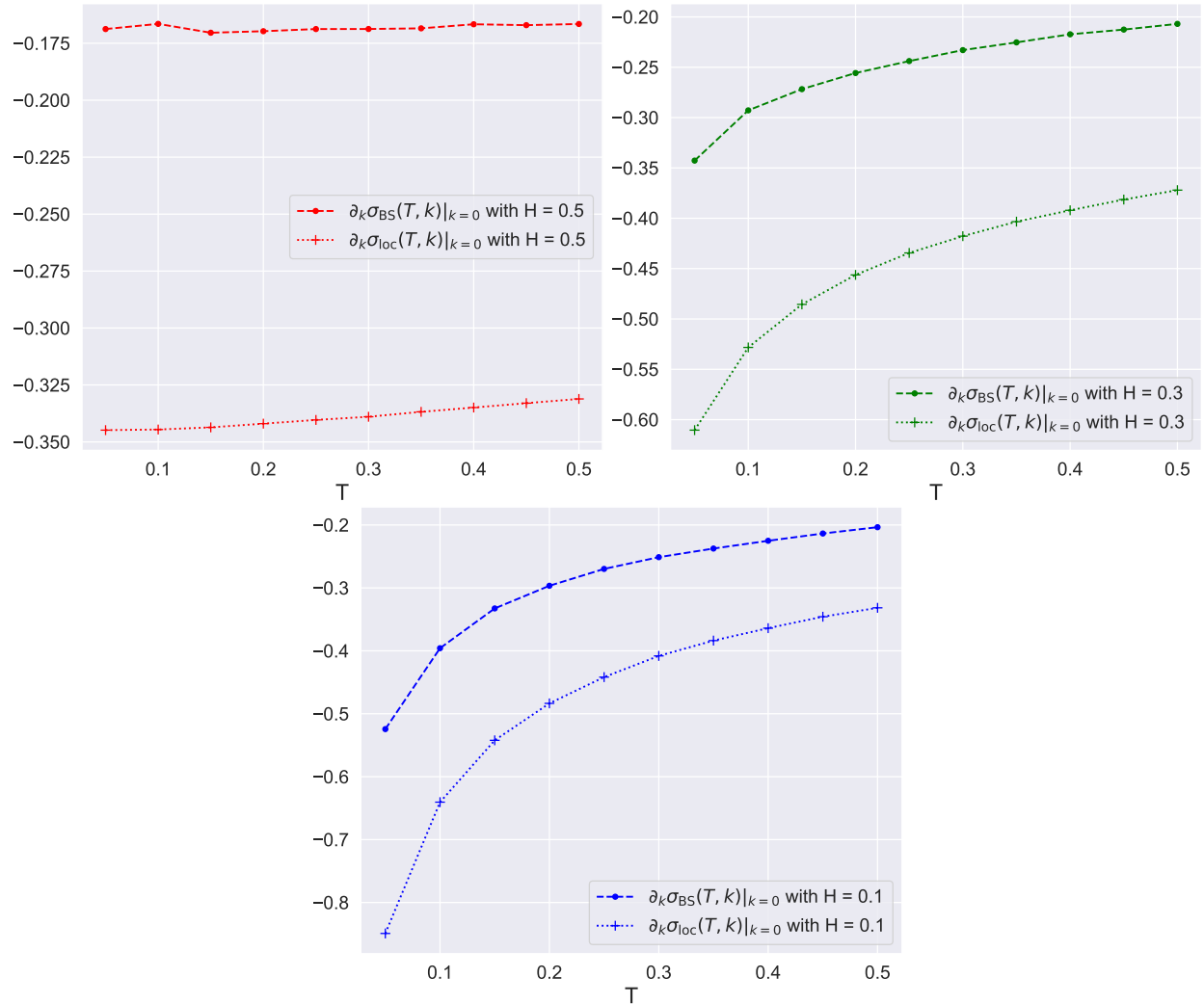

Figure 1: At-the-money implied and local volatility skews in the rough Bergomi model (4.1) for $H = 0.5$ (red, top left figure), $H = 0.3$ (green, top right figure), and $H = 0.1$ (blue, bottom figure). The maturity T on the x-axis is expressed in years.
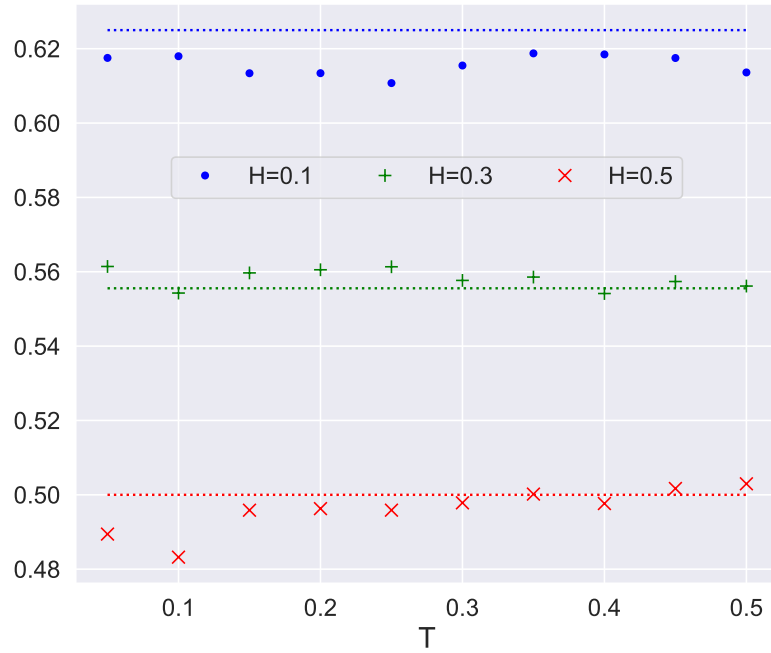
Figure 2: Numerical evidence for the $\frac{1}{H+3/2}$ ratio rule stated in Corollary 3.4: we plot the ratio of the at-the-money implied and local volatility skews $\frac{\partial_k \sigma_{\mathrm{BS}}(T,k)|_{k=0}}{\partial_k \sigma_{\mathrm{loc}}(T,k)|_{k=0}}$ for $H \in \{0.1, 0.3, 0.5\}$ against maturity T (in years). The dashed lines correspond to the constant values $\frac{1}{H+3/2}$ (blue for $H = 0.1$, green for $H = 0.3$, red for $H = 0.5$).
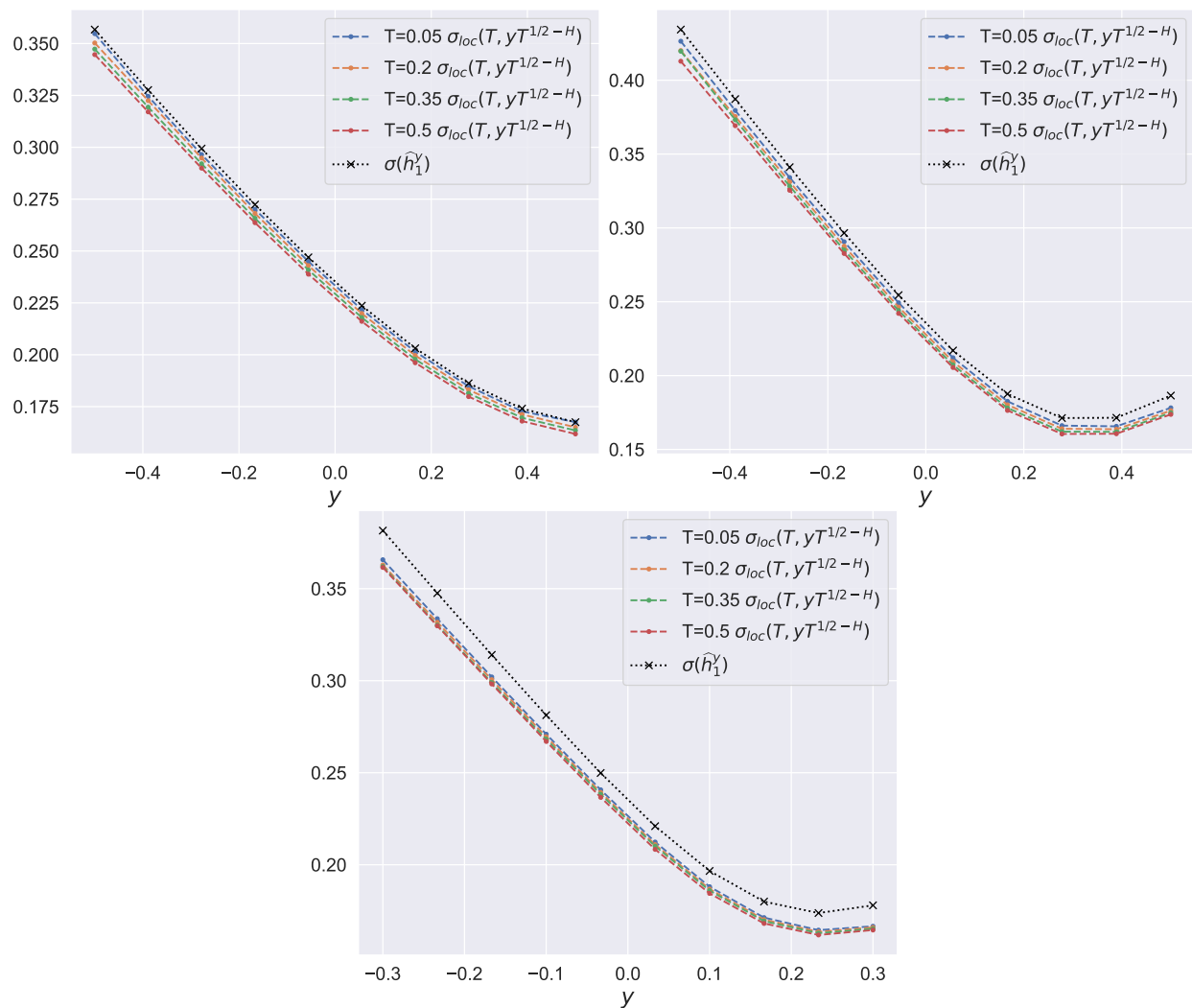
Figure 3: Short-dated local volatility the rough Bergomi model (4.1) for $H = 0.5$ (top left figure), $H = 0.3$ (top right figure), and $H = 0.1$ (bottom figure). Recall that, according to Theorem 3.3, $\sigma_{\mathrm{loc}}(T, y\,T^{1/2-H}) \to \sigma(\widehat{h}_1^y)$ as $T \to 0$. The rate function minimizing path $\widehat{h}_t^y$ is evaluated using the Ritz projection method with $N = 8$ Fourier basis functions, see section 4.2.
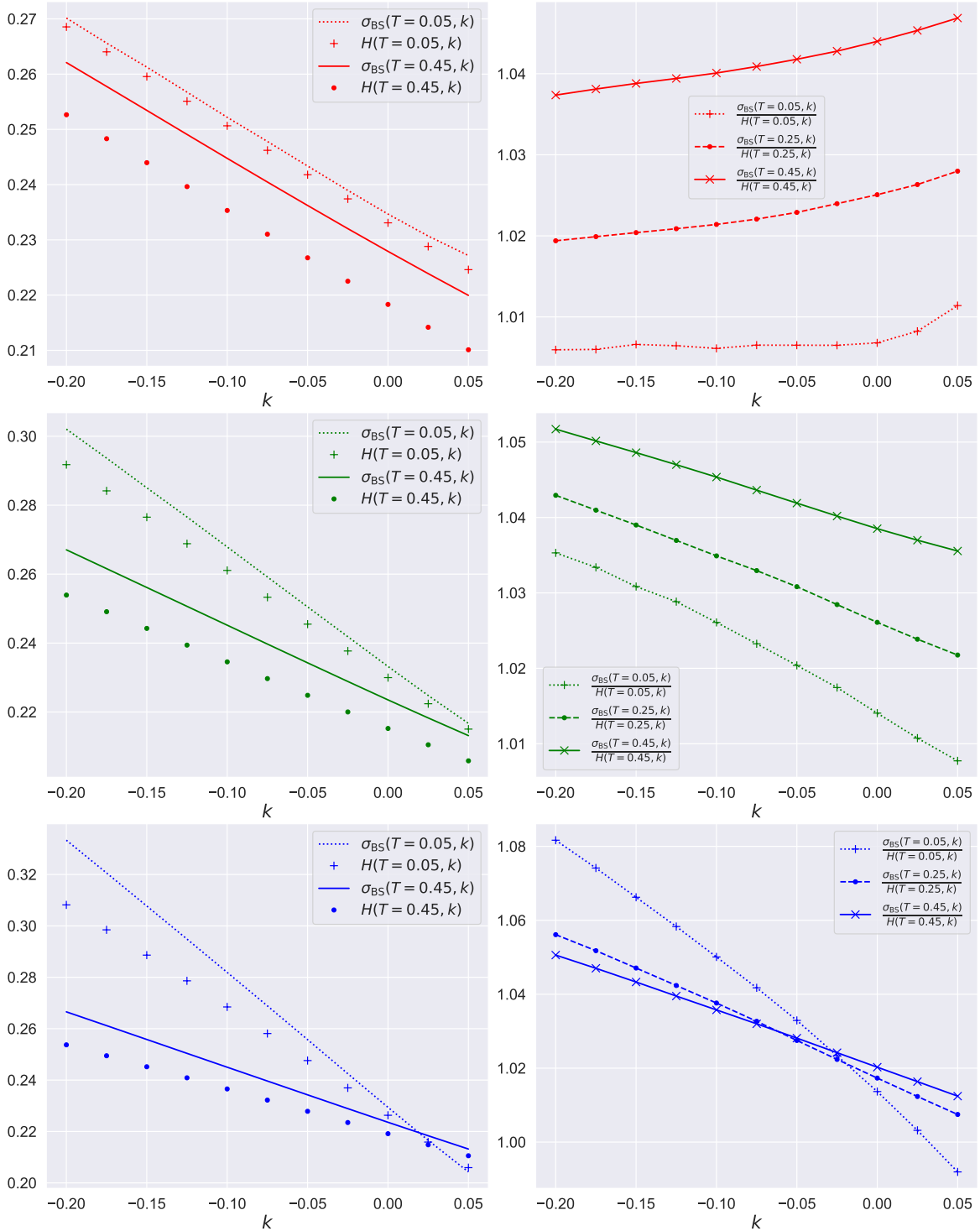
Figure 4: Numerical evidence for the failure of the harmonic mean formula within the rough Bergomi model (4.1) (see Remark 3.6): in the left figures, we compare the implied volatility $\sigma_{\mathrm{BS}}(T, k)$ and the harmonic mean $H(T, k)$ of the local volatility defined in (3.16), for two different maturities T and for $H = 0.5$ (red), $H = 0.3$ (green), and $H = 0.1$ (blue). In the right figures (same color conventions as the left figures), we plot the ratio $\frac{\sigma_{\mathrm{BS}}(T,k)}{H(T,k)}$ of the two functions, expected to tend to 1 as $T \to 0$ when $H = 0.5$.

15

5 Proofs

Proof of Corollary 3.4. Equation (3.12) is a straightforward consequence of Theorem 3.3.

It is a standard result that the implied volatility σ_{BS} and the local volatility σ_{loc} generated by a stochastic volatility model with $\rho = 0$ are symmetric around $y = 0$, so that the finite-difference at-the-money skews $\mathcal{S}_{\mathrm{BS}}$ and $\mathcal{S}_{\mathrm{loc}}$ are identically zero in this case. We, therefore, assume $\rho \neq 0$ in what follows. Let us write $\langle K1, 1 \rangle = \int_0^1 K1(t)\mathrm{d}t$ and $K1(t) = \int_0^t K(t,s)\mathrm{d}s$. Using an expansion of the map $y \mapsto \widehat{h}_1^y$ around $y = 0$ as provided in [25], we have

$$\Sigma(y) = \sigma\left(\widehat{h}_1^y\right) = \sigma_0 + y\frac{\sigma_0'}{\sigma_0}\rho K1(1) + \mathcal{O}(y^2) \qquad \text{as } y \to 0\,.$$

Together with (3.12), this implies

$$\mathcal{S}_{\mathrm{loc}}(t, y) \sim \left(\frac{\sigma_0'}{\sigma_0}\rho K1(1) + r(y)\right)\frac{1}{t^{1/2-H}}$$

as $t \to 0$, where $r(y) \to 0$ as $y \to 0$. Similarly, from (3.7) and a third-order energy expansion of Λ, obtained in [6, Thm 3.4] (an extension to forth order is given in [25] but not required here) it follows that

$$\mathcal{S}_{\mathrm{BS}}(t, y) \sim \frac{\chi(y) - \chi(-y)}{2y}\frac{1}{t^{1/2-H}} = \left(\frac{\sigma_0'}{\sigma_0}\rho\langle K1, 1\rangle + \ell(y)\right)\frac{1}{t^{1/2-H}}$$

as $t \to 0$, where $\ell(y) \to 0$ as $y \to 0$. Therefore

$$\frac{\mathcal{S}_{\mathrm{BS}}(t, y)}{\mathcal{S}_{\mathrm{loc}}(t, y)} \to \frac{\chi(y) - \chi(-y)}{\Sigma(y) - \Sigma(-y)} = \frac{\langle K1, 1\rangle}{K1(1)} + o(1) \qquad \text{as } t \to 0.$$

The identity

$$K1(1) = (H + 3/2)\langle K1, 1\rangle$$

for $K(t, s) = \sqrt{2H}(t - s)^{H-1/2}$ is straightforward to prove using simple integration (we note in passing that this identity holds for any self-similar $\widehat{W}$, by leveraging a representation in [46]). The statement of the corollary follows. $\qquad\square$

The proof of Theorem 3.3 is based on the following representation of the Markovian projection, based on the integration by parts of the Malliavin calculus,

$$\mathbb{E}[V_T | X_T = y] = \frac{\mathbb{E}\left[V_T \mathbf{1}_{X_T \geq y}\int_0^T \frac{1}{\overline{\rho}\sigma(\widehat{W}_t)}\mathrm{d}\overline{W}_t\right]}{\mathbb{E}\left[\mathbf{1}_{X_T \geq y}\int_0^T \frac{1}{\overline{\rho}\sigma(\widehat{W}_t)}\mathrm{d}\overline{W}_t\right]}. \tag{5.1}$$

Representation (5.1) is rather classical, see [23], though spelled out only in case $\rho = 0$, and [3] for a general formula in an abstract setting. We note that $\overline{\rho}$ cancels as long as it is not zero, equivalently $|\rho| < 1$, which is our non-degeneracy assumption. (We kept $\overline{\rho}$ above to insist on this point.) For completeness, we give a proof in Lemma 5.1.

Proof of Theorem 3.3. Setting $T = \varepsilon^2$ and using the time-scaling property of the triple $(\overline{W}, \widehat{W}, X)$, we get

$$\mathbb{E}\Big[\sigma(\widehat{W}_T)^2 \mathbf{1}_{X_T \geq yT^{1/2-H}} \int_0^T \frac{1}{\overline{\rho}\sigma(\widehat{W}_t)} \mathrm{d}\overline{W}_t\Big] = \mathbb{E}\Big[\sigma(\varepsilon^{2H}\widehat{W}_1)^2 \mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}} \int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)} \varepsilon \mathrm{d}\overline{W}_t\Big]$$

$$\mathbb{E}\Big[\mathbf{1}_{X_T \geq yT^{1/2-H}} \int_0^T \frac{1}{\overline{\rho}\sigma(\widehat{W}_t)} \mathrm{d}\overline{W}_t\Big] = \mathbb{E}\Big[\mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}} \int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)} \varepsilon \mathrm{d}\overline{W}_t\Big].$$

We define

$$J(\varepsilon, y) = e^{\frac{\Lambda(y)}{\varepsilon^{4H}}} \mathbb{E}\Big[\sigma(\varepsilon^{2H}\widehat{W}_1)^2 \mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}} \int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)} \varepsilon \mathrm{d}\overline{W}_t\Big]$$

$$\overline{J}(\varepsilon, y) = e^{\frac{\Lambda(y)}{\varepsilon^{4H}}} \mathbb{E}\Big[\mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}} \int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)} \varepsilon \mathrm{d}\overline{W}_t\Big] \tag{5.2}$$

so that $\sigma^2_{\mathrm{loc}}(\varepsilon^2, y\,\varepsilon^{1-2H}) = \frac{J(\varepsilon, y)}{\overline{J}(\varepsilon, y)}$ from (5.1). The implementation of an infinite-dimensional Laplace method along the lines of [24] allows us to determine the asymptotic behavior of $J(\varepsilon, y)$ and $\overline{J}(\varepsilon, y)$ as $\varepsilon \to 0$: we postpone the details to Lemma 5.2 below. We obtain the $\varepsilon \to 0$ limit of $\sigma^2_{\mathrm{loc}}(\varepsilon^2, y\,\varepsilon^{1-2H})$, and therefore the statement of the Theorem, from (5.3). $\qquad\square$

Lemma 5.1. *The representation formula* (5.1) *for the conditional expectation holds for every* $y \in \mathbb{R}$.

Proof. The Malliavin derivative $\overline{D}$ of X_T with respect to $\overline{W}$ is

$$\overline{D}_t X_T = \overline{\rho}\sqrt{V}_t, \qquad t < T,$$

because V is W-adapted. Consider a two-dimensional Skorohod integrable process $(0, \overline{u})$, with

$$\overline{u}_t = \frac{1}{T\overline{D}_t X_T} = \frac{1}{T\overline{\rho}\sqrt{V}_t}.$$

We write ∂ for the Skorohod integral. From Malliavin integration by parts formula for a bounded smooth function $\phi : \mathbb{R} \to \mathbb{R}$ and using that $\overline{D}_t V_T = 0$, we obtain

$$\mathbb{E}[V_T\phi(X_T)\partial(0, \overline{u})] = \mathbb{E}[\langle\overline{D}(V_T\phi(X_T)), \overline{u}\rangle] = \mathbb{E}[V_T\phi'(X_T)\langle\overline{D}X_T, \overline{u}\rangle]$$

$$= \mathbb{E}\Big[V_T\phi'(X_T) \int_0^T \overline{D}_t X_T\,\overline{u}_t \mathrm{d}t\Big] = \mathbb{E}[V_T\phi'(X_T)].$$

We have used here boundedness of $\phi(\cdot)$ and Assumptions 2.1 on $\sigma(\cdot)$ (see proof of next Lemma 5.4 for a detailed argument). Moreover, since V is adapted, we have $\partial(0, \overline{u}) = \int_0^T \frac{1}{T\overline{\rho}\sqrt{V}_t} \mathrm{d}\overline{W}_t$, and so

$$\mathbb{E}[V_T\phi'(X_T)] = \frac{1}{T\overline{\rho}}\mathbb{E}\Big[V_T\phi(X_T) \int_0^T \frac{1}{\sqrt{V}_t} \mathrm{d}\overline{W}_t\Big]$$

Following the same steps, one can show that the following identity also holds:

$$\mathbb{E}[\phi'(X_T)] = \frac{1}{\overline{\rho}T}\mathbb{E}\Big[\phi(X_T)\Big(\int_0^T \frac{1}{\sqrt{V}_t} \mathrm{d}\overline{W}_t\Big)\Big].$$

The representation formula (5.1) for the conditional expectation then follows from a regularization procedure of the indicator function $\mathbf{1}_{X_t \geq y}$, see for example [3]. $\qquad\square$

17

Lemma 5.2. *For $y \in \mathbb{R} \setminus \{0\}$ small enough, one has $\int_0^1 \frac{\mathrm{d}\overline{h}_t^y}{\sigma(\widehat{h}_t^y)} \neq 0$. Moreover, for $J, \overline{J}$ defined in* (5.2) *we have*

$$
\begin{aligned}
J(\varepsilon, y) &\sim \varepsilon^{2H} \sigma^2(\widehat{h}_1^y) \left(\int_0^1 \frac{\mathrm{d}\overline{h}_t^y}{\sigma(\widehat{h}_t^y)} \right) \frac{1}{\overline{\rho}\sqrt{2\pi}\sqrt{2\Lambda(y)}} \mathbb{E}\left[\exp\left(\Lambda'(y) \Delta_2 \right) \right], \\
\overline{J}(\varepsilon, y) &\sim \varepsilon^{2H} \left(\int_0^1 \frac{\mathrm{d}\overline{h}_t^y}{\sigma(\widehat{h}_t^y)} \right) \frac{1}{\overline{\rho}\sqrt{2\pi}\sqrt{2\Lambda(y)}} \mathbb{E}\left[\exp\left(\Lambda'(y) \Delta_2 \right) \right]
\end{aligned}
\tag{5.3}
$$

as $\varepsilon \to 0$, where Δ_2 is a quadratic Wiener functional given in (5.8) *(see also [24, Equation (7.4)]).*

Corollary 5.3 (Digital expansion)**.** *We do not use it here but we note that from the computations in the proof of Theorem 3.3 and Lemma 5.2, it follows that, with $y > 0$ small enough,*

$$
\mathbb{P}\left(X_T \geq yT^{1/2 - H} \right) \sim e^{-\frac{\Lambda(y)}{T}} T^H \frac{1}{\sqrt{2\pi}\sqrt{2\Lambda(y)}} \mathbb{E}\left[\exp\left(\Lambda'(y) \Delta_2 \right) \right], \qquad as\ T \to 0.
$$

Proof of Lemma 5.2. We aim to apply the asymptotic results in [24]. Assumption (A1) in [24] is nothing but the validity of the large deviations principle for the model defined in (2.1), which we have already discussed in Remark 3.1. We take y close enough to 0 so that the non-degeneracy assumptions [24, Assumptions (A3), (A4), (A5)] are satisfied for the model under consideration, as it has been checked in [24, Section 7.1]. Therefore, we have that the preliminary regularity structures results in [24] apply to our setting, and can employ them in the proof.

Using [25, Proof of Lemma 3.4, Step 1] we have

$$
\int_0^1 g_t \mathrm{d}\overline{h}_t^y = \overline{\rho}\Lambda'(y) \int_0^1 g_t \sigma(\widehat{h}_t^y) \mathrm{d}t,
$$

for any square-integrable test function g, so that in particular

$$
\int_0^1 \frac{\mathrm{d}\overline{h}_t^y}{\sigma(\widehat{h}_t^y)} = \overline{\rho}\Lambda'(y) \int_0^1 1 \mathrm{d}t = \overline{\rho}\Lambda'(y)
$$

and $\Lambda'(y) \neq 0$, as detailed in the proof of [24, Theorem 6.1]. The proof of (5.3) is then a modification of [24, Proposition 8.7], from which we borrow the notations. Necessary definitions are recalled in Appendix A. We only prove the statement for J, the one for $\overline{J}$ being completely analogous. Set, for any $\delta > 0$,

$$
\mathbb{P}_\delta(A) = \mathbb{P}(A \cap \{\varepsilon^{2H}|||\mathbf{W}||| < \delta\}),
\tag{5.4}
$$

with $\mathbf{W}$ defined in (A.2), and set

$$
J_\delta(\varepsilon, y) = \varepsilon^{1-2H} e^{\frac{I(y)}{\varepsilon^{4H}}} \mathbb{E}_\delta\left[\sigma(\varepsilon^{2H}\widehat{W}_1)^2 \mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}} \left(\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)} \varepsilon^{2H} \mathrm{d}\overline{W}_t \right) \right]
$$

where the expectation $\mathbb{E}_\delta$ is with respect to the sub-probability $\mathbb{P}_\delta$. As a consequence of Lemma 5.4, any "algebraic expansion" of J (i.e., in powers of ε) does not change by switching to J_δ. So, proving the asymptotic behavior (5.3) for J_δ implies the statement.

18

We recall (3.1) and apply Girsanov's theorem, via the transformation

$$\varepsilon^{2H}\mathrm{W} \to \varepsilon^{2H}\mathrm{W} + \mathrm{h}^y = \varepsilon^{2H}(\mathrm{W} + \mathrm{h}^y/\varepsilon^{2H}),$$
$$\varepsilon^{2H}\widehat{W} \to \varepsilon^{2H}\widehat{W} + \widehat{h}^y = \varepsilon^{2H}(\widehat{W} + \widehat{h}^y/\varepsilon^{2H})$$

$$(5.5)$$

from which we introduce

$$\overline{Z}_1^\varepsilon = \int_0^1 \sigma\left(\varepsilon^{2H}\widehat{W}_t + \widehat{h}_t^y\right)\mathrm{d}[\varepsilon^{2H}\widetilde{W} + \widetilde{h}^y]_t - \frac{\varepsilon^{1+2H}}{2}\int_0^1 \sigma^2\left(\varepsilon^{2H}\widehat{W}_t + \widehat{h}_t^y\right)\mathrm{d}t, \qquad (5.6)$$

with stochastic Taylor expansion (A.4). From Girsanov theorem, $J_\delta(\varepsilon,y)/\varepsilon^{1-2H}$ equals

$$= e^{\frac{1}{2\varepsilon^{4H}}\|\mathrm{h}^y\|_{H^1}^2}\mathbb{E}_\delta\left[\sigma(\varepsilon^{2H}\widehat{W}_1)^2\mathbf{1}_{\varepsilon^{2H-1}X_1^\varepsilon \geq y}\left(\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)}\varepsilon^{2H}\mathrm{d}\overline{W}_t\right)\right]$$

$$= \mathbb{E}_\delta\left[e^{-\frac{1}{\varepsilon^{2H}}\int_0^1 \dot{h}^y\mathrm{d}W}\sigma^2(\varepsilon^{2H}\widehat{W}_1 + \widehat{h}_1^y)\mathbf{1}_{\varepsilon^{2H}g_1 + \varepsilon^{4H}g_2 + r_3 \geq 0}\left(\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t + \widehat{h}_t^y)}(\varepsilon^{2H}\mathrm{d}\overline{W}_t + \mathrm{d}\overline{h}_t^y)\right)\right].$$

Theorem A.1, applied with ε^{2H} (instead of ε), gives on $\{\varepsilon^{2H}|||\mathbf{W}||| < \delta\}$,

$$\sigma^2(\varepsilon^{2H}\widehat{W}_1 + \widehat{h}_1^y) = \sigma^2(\widehat{h}_1^y) + \ell_{\varepsilon,\mathbf{W}}^1,$$

with $|\ell_{\varepsilon,\mathbf{W}}^1| \leq C\delta$ and

$$\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t + \widehat{h}_t^y)}(\varepsilon^{2H}\mathrm{d}\overline{W}_t + \mathrm{d}\overline{h}_t^y) = \int_0^1 \frac{1}{\overline{\rho}\sigma(\widehat{h}_t^y)}\mathrm{d}\overline{h}_t^y + \ell_{\varepsilon,\mathbf{W}}^2$$

with $|\ell_{\varepsilon,\mathbf{W}}^2| \leq C\varepsilon^{2H}|||\mathbf{W}||| \leq C\delta$. Therefore,

$$\sigma^2(\varepsilon^{2H}\widehat{W}_1 + \widehat{h}_1^y)\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t + \widehat{h}_t^y)}(\varepsilon^{2H}\mathrm{d}\overline{W}_t + \mathrm{d}\overline{h}_t^y) = \sigma^2(\widehat{h}_1^y)\int_0^1 \frac{1}{\overline{\rho}\sigma(\widehat{h}_t^y)}\mathrm{d}\overline{h}_t^y + \ell_{\varepsilon,\mathbf{W}}$$

with $|\ell_{\varepsilon,\mathbf{W}}| \leq C\delta$. If $\varepsilon^{2H}|||\mathbf{W}||| \leq \delta$ we also have (A.7), so, for fixed δ, for ε small enough,

$$|r_3^\varepsilon| \leq \delta\varepsilon^{4H}(C + |||\mathbf{W}|||^2).$$

We have

$$\mathbb{E}_\delta[\dots] \in \left(\sigma^2(\widehat{h}_1)\int_0^1 \frac{\mathrm{d}\overline{h}_t^y}{\overline{\rho}\sigma(\widehat{h}_t^y)} \pm C\delta\right)\mathbb{E}_\delta\left[e^{-\frac{1}{\varepsilon^{2H}}\int_0^1 \dot{h}^y\mathrm{d}W}\mathbf{1}_{(g_1 + \varepsilon^{2H}g_2 \pm \delta\varepsilon^{2H})(C + |||\mathbf{W}|||^2) \geq 0}\right] \qquad (5.7)$$

The optimal condition [24, Lemma C.3] gives $\int_0^1 \dot{h}^y\mathrm{d}W = \Lambda'(y)g_1$. By [24, Lemma 8.3],

$$g_2 = \Delta_2 + g_1\Delta_1 + g_1^2\Delta_0 \qquad (5.8)$$

where the Δ_i's are independent of g_1. We set now, as in [2], the zero-mean Gaussian process $\mathrm{V} = \mathrm{V}^y$

$$\mathrm{V}_t(\omega) := \mathrm{W}_t(\omega) - g_1(\omega)\mathrm{v}_t \qquad (5.9)$$

where v is chosen so that V is independent of g_1. We also define

$$\mathbf{V}(\omega) := T_{-g_1(\omega)\text{v}}\mathbf{W}(\omega)$$

where T, the "lifted" Cameron–Martin translation, is defined in (A.3). We let

$$\widetilde{\Delta}_0 := \Delta_0 + C\delta\|\text{v}\|_{H^1}^2,$$
$$\widetilde{\Delta}_2^{\pm} := \Delta_2 \pm \delta(C + \||\mathbf{V}\||^2),$$

where $\widetilde{\Delta}_2^{\pm}$ is also P-independent of g_1 and $\mathbf{V}$. We refer to [24] for details, and here we only use that $\varepsilon^{2H}\||\mathbf{V}\|| \le C\delta$, so that

$$|\varepsilon^{2H}\Delta_1| \le C\varepsilon^{2H}\||\mathbf{V}\|| \le C\delta,$$

when $\varepsilon^{2H}\||\mathbf{W}\|| \le \delta$. Thus, the asymptotic behavior of $J_\delta(\varepsilon, y)$ is sandwiched by $\sigma^2(\widehat{h}_1^y) \int_0^1 \frac{\mathrm{d}\overline{h}_t^y}{\overline{\rho}\sigma(\widehat{h}_t^y)} \pm$ $C\delta$ times $(*)$ with

$$(*) \in \mathbb{E}_\delta\left[\exp\left(-\frac{\Lambda'(y)g_1}{\varepsilon^{2H}}\right)\mathbf{1}_{g_1 + \varepsilon^{2H}\widetilde{\Delta}_2^{\pm} \pm C(1+\widetilde{\Delta}_0)\delta|g_1| > 0}\right].$$

The limit of this expectation can be computed with the Laplace method. We prove the upper bound. Clearly,

$$\mathbb{E}_\delta\left[\exp\left(-\frac{\Lambda'(y)g_1}{\varepsilon^{2H}}\right)\mathbf{1}_{g_1 + \varepsilon^{2H}\widetilde{\Delta}_2^{\pm} \pm C(1+\widetilde{\Delta}_0)\delta|g_1| > 0}\right] \le \mathbb{E}\left[\cdots\right]$$

where $\cdots$ means the same argument. Set $\sigma_y = \sqrt{2\Lambda(y)}/\Lambda'(y)$ and

$$\gamma_\delta := C(1 + \widetilde{\Delta}_0)\delta/\sigma_y$$

and assume that δ is small enough that $\gamma_\delta < 1$. By [24, Theorem 6.1, part (iii)], we have $\frac{\varepsilon^{2H}}{\Lambda'(y)\sigma_y} > 0$ and can then apply Lemma 5.5 (with $N = g_1/\sigma_y$) to see that

$$\mathbb{E}\left[\cdots|\Delta_2, \mathbf{V}\right] \le \frac{\varepsilon^{2H}}{\sigma_y\Lambda'(y)\sqrt{2\pi}} \max\left[e^{\frac{I'(y)\left(\Delta_2 + \delta\left(C + \||\mathbf{V}\||^2\right)\right)}{1-\gamma_\delta}}, e^{\frac{I'(y)\left(\Delta_2 + \delta\left(C + \||\mathbf{V}\||^2\right)\right)}{1+\gamma_\delta}}\right].$$

By [24, Proposition 8.6 and proof of Corollary 7.1], $\exp\left(\Lambda'(y)\Delta_2\right) \in L^{1+}$ and by [24, Lemma 8.3 (iv)] $\exp(\||\mathbf{V}\||^2) \in L^{0+}$, so that by letting successively ε and δ go to 0 we obtain that

$$\limsup_{\varepsilon \to 0} \varepsilon^{-2H}\mathbb{E}\left[\cdots\right] \le \frac{1}{\sigma_y\Lambda'(y)\sqrt{2\pi}}\mathbb{E}\left[\exp\left(\Lambda'(y)\Delta_2\right)\right].$$

Recalling now $\sqrt{2\Lambda(y)} = \Lambda'(y)\sigma_y$, and the prefactor $\sigma^2(\widehat{h}_1^y) \int_0^1 \frac{\mathrm{d}\overline{h}_t^y}{\overline{\rho}\sigma(\widehat{h}_t^y)} \pm C\delta$ in (5.7) we have the upper bound. The lower bound is proved in the same way using the lower bound in Lemma 5.5. $\square$

Lemma 5.4. *Fix $\delta > 0$. Then there exists $c = c_{y,\delta} > 0$ such that*

$$|J_\delta(\varepsilon, y) - J(\varepsilon, y)| = \mathcal{O}(\exp(-c/\overline{\varepsilon}^2)).$$

Proof. To this end, recall the sub-probability (5.4) and introduce

$$B = \{\varepsilon^{2H}|||\mathbf{W}||| \geq \delta\}^c.$$

We have

$$J(\varepsilon, y) - J_\delta(\varepsilon, y) = \exp\left(\frac{I(y)}{\varepsilon^{4H}}\right) \mathbb{E}\left[\sigma(\varepsilon^{2H}\widehat{W}_1)^2 \mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}}\left(\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}W_t^H)}\varepsilon\mathrm{d}\overline{W}_t\right)\mathbf{1}_B\right]. \quad (5.10)$$

We have, for any $p, p' > 1$ conjugate exponents,

$$\mathbb{E}\left|\sigma(\varepsilon^{2H}\widehat{W}_1)^2 \mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}}\left(\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)}\varepsilon\mathrm{d}\overline{W}_t\right)\mathbf{1}_B\right|$$

$$\leq \varepsilon\left[\mathbb{E}\left|\sigma(\varepsilon^{2H}\widehat{W}_1)^2\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)}\mathrm{d}\overline{W}_t\right|^p\right]^{1/p} \mathbb{E}[\mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}}\mathbf{1}_B]^{1/p'}. \quad (5.11)$$

The second factor can be bounded using Hölder inequality as

$$\mathbb{E}\left|\sigma(\varepsilon^{2H}\widehat{W}_1)^2\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)}\mathrm{d}\overline{W}_t\right|^p \leq (\mathbb{E}[\sigma(\varepsilon^{2H}\widehat{W}_1)^{2q}])^{1/q}\left(\mathbb{E}\left|\int_0^1 \frac{1}{\overline{\rho}\sigma(\varepsilon^{2H}\widehat{W}_t)}\mathrm{d}\overline{W}_t\right|^{pq'}\right)^{1/q'}.$$

Since $\sigma(\cdot)$ satisfies (2.3), the first factor is bounded for any $q > 1$. Using Burkholder–Davis–Gundy inequality,

$$\mathbb{E}\left|\int_0^1 \frac{1}{\sigma(\varepsilon^{2H}\widehat{W}_t)}\mathrm{d}\overline{W}_t\right|^{pq'} \leq \mathbb{E}\left|\int_0^1 \frac{\mathrm{d}t}{\sigma(\varepsilon^{2H}\widehat{W}_t)^2}\right|^{pq'/2}$$

using condition (2.2) and the moment formula for log-normal variables,

$$\cdots \leq \mathbb{E}\left|\int_0^1 \exp(c\varepsilon^{2H}\widehat{W}_t)\mathrm{d}t\right|^{pq'/2} \leq \mathbb{E}\left|\exp(c\varepsilon^{2H}\widehat{W}_1)\right|^{pq'/2} < \infty$$

We conclude that the second factor in (5.11) is bounded by a constant, for any $p \geq 1$. In [24, lines after (8.5)] it is shown that

$$\mathbb{E}[\mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}}\mathbf{1}_B] = \mathcal{O}(e^{-\frac{\Lambda(\delta,y)}{\varepsilon^{4H}}})$$

with $\Lambda(\delta, y) > \Lambda(y)$. Now,

$$\mathbb{E}[\mathbf{1}_{X_1^\varepsilon \geq y\varepsilon^{1-2H}}\mathbf{1}_B]^{1/p'} = \mathcal{O}(e^{-\frac{\Lambda(\delta,y)}{p'\varepsilon^{4H}}})$$

and we can choose $p' > 1$ close enough to 1 in order to have $\Lambda(\delta, y)/p' > \Lambda(y)$. The statement follows. $\square$

Lemma 5.5. *Let $\alpha \in \mathbb{R}$, $\gamma \in [0, 1)$, $\varepsilon > 0$, and $N \sim \mathcal{N}(0, 1)$. Then for some $C > 0$, it holds that*

$$\min\left[e^{\frac{\alpha}{1-\gamma}}, e^{\frac{\alpha}{1+\gamma}}\right] - \varepsilon^2 \max\left[e^{\frac{\alpha}{1-\gamma}}\frac{\alpha^2 - 2\alpha(1-\gamma) + 2(1-\gamma)^2}{(1-\gamma)^2}, e^{\frac{\alpha}{1+\gamma}}\frac{\alpha^2 - 2\alpha(1+\gamma) + 2(1+\gamma)^2}{(1+\gamma)^2}\right]$$

$$(5.12)$$

$$\leq \sqrt{2\pi}\varepsilon^{-1}\mathbb{E}\left[\exp\left(-\varepsilon^{-1}N\right)\mathbf{1}_{N+\gamma|N|+\varepsilon\alpha>0}\right] \quad (5.13)$$

$$\leq \min\left[e^{\frac{\alpha}{1-\gamma}}, e^{\frac{\alpha}{1+\gamma}}\right] \quad (5.14)$$

Proof. The middle expression (5.12) equals

$$\varepsilon^{-1} \int_{-\infty}^{+\infty} e^{-y^2/2} e^{-\frac{y}{\varepsilon}} \mathbf{1}_{y+\gamma|y|+\varepsilon\alpha>0} \mathrm{d}y = \int_{-\infty}^{+\infty} e^{-v^2\varepsilon^2/2} e^{-v} \mathbf{1}_{v+\gamma|v|+\alpha>0} \mathrm{d}v.$$

Inequalities $1 - y^2/2 \leq \exp(-y^2/2) \leq 1$ lead to the stated bounds. Indeed,

$$(5.12) \leq \int_{-\infty}^{+\infty} e^{-v} \mathbf{1}_{v+\gamma|v|+\alpha>0} \mathrm{d}v$$

which is computable. The right-hand side equals

$$\int_{-\frac{\alpha}{1+\gamma}}^{\infty} e^{-v} dv = e^{\frac{\alpha}{1+\gamma}}, \text{ when } \alpha < 0$$

$$\int_{-\frac{\alpha}{1-\gamma}}^{\infty} e^{-v} \mathrm{d}v = e^{\frac{\alpha}{1-\gamma}}, \text{ when } \alpha \geq 0.$$

To obtain the lower bound, use $e^{-y^2/2} \geq 1 - y^2/2$ and split the integral to obtain

$$(5.12) \geq \int_{-\infty}^{+\infty} e^{-v} \mathbf{1}_{v+\gamma|v|+\alpha>0} \mathrm{d}v - \frac{\varepsilon^2}{2} \int_{-\infty}^{\infty} e^{-v} v^2 \mathbf{1}_{v+\gamma|v|+\alpha>0} \mathrm{d}v.$$

The first integral is computed as before. For the second one, when $\alpha < 0$ we have

$$e^{\frac{\alpha}{1+\gamma}} \frac{\alpha^2 - 2\alpha(1+\gamma) + 2(1+\gamma)^2}{(1+\gamma)^2}$$

while when $\alpha \geq 0$,

$$e^{\frac{\alpha}{1-\gamma}} \frac{\alpha^2 - 2\alpha(1-\gamma) + 2(1-\gamma)^2}{(1-\gamma)^2}.$$

The statement follows. $\qquad\square$

A   Elements of regularity structures for rough volatility

This appendix is based on [5, 24]. We have an fBm $\widehat{W} = K^H * \dot{W}$ of Hurst parameter H. Let M be the smallest integer such that $(M+1)H - 1/2 > 0$ and then pick κ small enough such that

$$(M+1)(H-\kappa) - 1/2 - \kappa > 0. \tag{A.1}$$

When $H = 1/2$, we have $M = 1$ and so $1/2 - \kappa \in (1/3, 1/2)$. This corresponds to the rough path case. More generally, we work with an enhancement of the Brownian noise $(W, \overline{W})$, also known as model, of the form

$$\mathbf{W}(\omega) = \left(W, \overline{W}, \widehat{W}, \int \widehat{W} \mathrm{d}W, \int \widehat{W} \mathrm{d}\overline{W}, \int \widehat{W}^2 \mathrm{d}W, \cdots, \int \widehat{W}^M \mathrm{d}\overline{W} \right), \tag{A.2}$$

with *homogeneous model norm*[2]

$$|||\mathbf{W}||| := \|W\|_{1/2-\kappa} + \|\overline{W}\|_{1/2-\kappa} + \|\widehat{W}\|_{H-\kappa} + \cdots + \|\int \widehat{W}^M d\overline{W}\|_{M(H-\kappa)+1/2-\kappa}^{1/3}$$

where $\|\cdot|_{1/2-\kappa}$ are classical, resp. 2-parameter, Hölder (semi)norms. One naturally defines, with $\mathrm{h} = (h, \overline{h}) \in H^1$ and $\widehat{h} = K^H * h$

$$T_{\mathrm{h}}(\mathbf{W}) = \left(W + h, \overline{W} + \overline{h}, \widehat{W} + \widehat{h}, \int (\widehat{W} + \widehat{h}) d(W + \overline{h}), \ldots \right). \tag{A.3}$$

Also, recall from [5] that there is a well-defined dilation δ_ε acting on models. Formally, it is obtained by replacing each occurrence of $W, \overline{W}, \widehat{W}$ with ε times that quantity:

$$\delta_\varepsilon \mathbf{W} = \left(\varepsilon W, \varepsilon \overline{W}, \varepsilon \widehat{W}, \varepsilon^2 \int \widehat{W} dW, \varepsilon^3 \int \widehat{W}^2 dW, \ldots \right) \in \mathcal{M},$$

where $\mathcal{M}$ is the space of models. As a consequence, dilation works well with homogeneous model norms,

$$|||\delta_\varepsilon \mathbf{W}||| = \varepsilon |||\mathbf{W}|||.$$

Theorem A.1 (Stochastic Taylor-like expansion)**.** *Let f be a smooth function. Fix $\mathrm{h} \in H^1$ and $\varepsilon > 0$. If $\mathbf{W}$ is a model, then so is $T_{\mathrm{h}}(\delta_\varepsilon \mathbf{W})$. The path-wise "rough/model" integral*

$$\Psi(\varepsilon) := \int_0^1 f\left(\varepsilon \widehat{W}_t + \widehat{h}_t \right) \mathrm{d}(T_{\mathrm{h}}(\delta_\varepsilon \mathbf{W}))_t$$

is well-defined, continuously differentiable in ε, and we have the estimates

$$|f(\varepsilon \widehat{W}_1 + \widehat{h}_1) - f(\widehat{h}_1)| = \mathcal{O}(\varepsilon |||\mathbf{W}|||),$$
$$|\Psi(\varepsilon) - \Psi(0)| = \mathcal{O}(\varepsilon |||\mathbf{W}|||),$$

valid on bounded sets of $\varepsilon |||\mathbf{W}|||$.

Proof. As in [24, Theorem B.6], just stop the expansion at the first order. $\square$

Lemma A.2. *Let $\overline{Z}_1^\varepsilon$ be defined in (5.6) and recall $\widehat{\varepsilon} \equiv \varepsilon^{2H}$. Then*

$$\overline{Z}_1^\varepsilon = g_0 + \widehat{\varepsilon} g_1(\omega) + \widehat{\varepsilon}^2 g_2(\omega) + r_3(\omega) \tag{A.4}$$

with $g_0 = y$,

$$g_1 = \int_0^1 \sigma'(\widehat{h}_s^y)\widehat{W}_s \mathrm{d}\widetilde{h}_s^y + \int_0^1 \sigma(\widehat{h}_s^y)\mathrm{d}\widetilde{W}_s, \tag{A.5}$$

$$g_2 = \frac{1}{2}\int_0^1 \sigma''(\widehat{h}_s^y)\widehat{W}_s^2 \mathrm{d}\widetilde{h}_s^y + \int_0^1 \sigma'(\widehat{h}_s^y)\widehat{W}_s \mathrm{d}\widetilde{W}_s \tag{A.6}$$

$$|r_3(\omega)| \le O(\varepsilon^{6H}|||\mathbf{W}|||^3) + O(\varepsilon^{1+2H}), \quad \textit{uniformly on bounded sets of } \varepsilon^{2H}|||\mathbf{W}|||. \tag{A.7}$$

---

[2]In fact, $\|W\|_{1/2-\kappa} \asymp \|\widehat{W}\|_{H-\kappa}$ by Schauder so that including $\widehat{W}$ is mildly redundant.

Proof. Directly from (5.6),

$$\overline{Z}_1^\varepsilon = \int_0^1 \sigma\left(\varepsilon^{2H}\widehat{W}_t + \widehat{h}_t^y\right) \mathrm{d}[\varepsilon^{2H}\widetilde{W} + \widetilde{h}^y]_t + O(\varepsilon^{1+2H})$$

uniformly on bounded sets of $\varepsilon^{2H}|\widetilde{W}|$, and hence on bounded sets of $\varepsilon^{2H}|||\mathbf{W}|||$. From [24, Theorem B.6], applied with ε replaced by ε^{2H}, and then again uniformly on bounded sets of $\varepsilon^{2H}|||\mathbf{W}|||$ we arrive at the error estimate,

$$|r_3(\omega)| \leq O(\varepsilon^{6H}|||\mathbf{W}|||^3) + O(\varepsilon^{1+2H}),$$

valid uniformly on bounded sets of $\varepsilon^{2H}|||\mathbf{W}|||$. $\qquad\square$

References

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

[2] R. Azencott. Petites perturbations aléatoires des systemes dynamiques: développements asymptotiques. *Bulletin des sciences mathématiques*, 109(3):253–308, 1985.

[3] V. Bally. An elementary introduction to Malliavin calculus. Research Report RR-4718, INRIA. Available at https://hal.inria.fr/inria-00071868., 2003.

[4] C. Bayer, P. Friz, and J. Gatheral. Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904, 2016.

[5] C. Bayer, P. K. Friz, P. Gassiat, J. Martin, and B. Stemper. A regularity structure for rough volatility. *Mathematical Finance*, 30(3):782–832, 2020.

[6] C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath, and B. Stemper. Short-time near-the-money skew in rough fractional volatility models. *Quantitative Finance*, 19(5):779–798, 2019.

[7] C. Bayer, C. B. Hammouda, and R. Tempone. Hierarchical adaptive sparse grids and quasi-Monte Carlo for option pricing under the rough Bergomi model. *Quantitative Finance*, 0(0):1–17, 2020.

[8] C. Bayer, F. A. Harang, and P. Pigato. Log-modulated rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 12(3):1257–1284, 2021.

[9] C. Bayer, B. Horvath, A. Muguruza, B. Stemper, and M. Tomas. On deep calibration of (rough) stochastic volatility models. *arXiv preprint arXiv:1908.08806*, 2019.

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

[11] M. Bennedsen, A. Lunde, and M. S. Pakkanen. Decoupling the short- and long-term behavior of stochastic volatility. *Journal of Financial Econometrics*, 2021.

[12] H. Berestycki, J. Busca, and I. Florent. Asymptotics and calibration of local volatility models. *Quantitative Finance*, 2:61–69, 2002.

[13] H. Berestycki, J. Busca, and I. Florent. Computing the implied volatility in stochastic volatility models. *Communications on Pure and Applied Mathematics*, 57(10):1352–1373, 2004.

[14] G. Brunick and S. Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4):1584 – 1628, 2013.

[15] S. De Marco, P. Friz, and S. Gerhold. Rational shapes of local volatility. *Risk*, 26(2):70, 2013.

[16] S. De Marco and P. K. Friz. Local Volatility, Conditioned Diffusions, and Varadhan's Formula. *SIAM Journal on Financial Mathematics*, 9(2):835–874, 2018.

[17] B. Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[18] B. Dupire. A unified theory of volatility. *Derivatives pricing: The classic collection, 2004 (P. Carr, ed.)*, pages 185–196, 1996.

[19] O. El Euch, M. Fukasawa, J. Gatheral, and M. Rosenbaum. Short-term at-the-money asymptotics under stochastic volatility models. *SIAM Journal on Financial Mathematics*, 10(2):491–511, 2019.

[20] O. El Euch, M. Fukasawa, and M. Rosenbaum. The microstructural foundations of leverage effect and rough volatility. *Finance and Stochastics*, 22(2):241–280, 2018.

[21] O. El Euch and M. Rosenbaum. The characteristic function of rough Heston models. *Mathematical Finance*, 29(1):3–38, 2019.

[22] M. Forde and H. Zhang. Asymptotics for rough stochastic volatility models. *SIAM Journal on Financial Mathematics*, 8(1):114–145, 2017.

[23] E. Fournié, J.-M. Lasry, J. Lebuchoux, and P.-L. Lions. Applications of Malliavin calculus to Monte-Carlo methods in finance. II. *Finance and Stochastics*, 5(2):201–236, 2001.

[24] P. K. Friz, P. Gassiat, and P. Pigato. Precise asymptotics: Robust stochastic volatility models. *The Annals of Applied Probability*, 31(2):896–940, 2021.

[25] P. K. Friz, P. Gassiat, and P. Pigato. Short-dated smile under rough volatility: asymptotics and numerics. *Quantitative Finance*, pages 1–18, 2021.

[26] P. K. Friz, J. Gatheral, A. Gulisashvili, A. Jacquier, and J. Teichmann. *Large deviations and asymptotic methods in finance*, volume 110. Springer, 2015.

[27] P. K. Friz, S. Gerhold, and M. Yor. How to make Dupires local volatility work with jumps. *Quantitative Finance*, 14(8):1327–1331, 2014.

[28] P. K. Friz and M. Hairer. *A Course on Rough Paths. With an introduction to regularity structures.* Springer, 2020.

[29] P. K. Friz, P. Pigato, and J. Seibel. The Step Stochastic Volatility Model. *Risk*, June, 2021.

[30] M. Fukasawa. Asymptotic analysis for stochastic volatility: martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.

[31] M. Fukasawa. Short-time at-the-money skew and rough fractional volatility. *Quantitative Finance*, 17(2):189–198, 2017.

[32] M. Fukasawa. Volatility has to be rough. *Quantitative Finance*, 21(1):1–8, 2021.

[33] M. Fukasawa, T. Takabatake, and R. Westphal. Is Volatility Rough? *arXiv preprint arXiv:1905.04852*, 2019.

[34] P. Gassiat. On the martingale property in the rough Bergomi model. *Electron. Commun. Probab.*, 24:9 pp., 2019.

[35] J. Gatheral. *The volatility surface: a practitioner's guide*. John Wiley & Sons, 2006.

[36] J. Gatheral, E. P. Hsu, P. Laurence, C. Ouyang, and T.-H. Wang. Asymptotics of implied volatility in local volatility models. *Mathematical Finance*, 22(4):591–620, 2012.

[37] J. Gatheral, T. Jaisson, and M. Rosenbaum. Volatility is rough. *Quantitative Finance*, pages 1–17, 2018.

[38] J. Gatheral and M. Keller-Ressel. Affine forward variance models. *Finance and Stochastics*, 23(3):501–533, 2019.

[39] L. Goudenège, A. Molent, and A. Zanette. Machine learning for pricing American options in high-dimensional Markovian and non-Markovian models. *Quantitative Finance*, 20(4):573–591, 2020.

[40] A. Gulisashvili. Large deviation principle for volterra type fractional stochastic volatility models. *SIAM Journal on Financial Mathematics*, 9(3):1102–1136, 2018.

[41] A. Gulisashvili. Gaussian stochastic volatility models: Scaling regimes, large deviations, and moment explosions. *Stochastic Processes and their Applications*, 130(6):3648 – 3686, 2020.

[42] J. Guyon and P. Henry-Labordère. Being particular about calibration. *Risk, January*, 2012.

[43] I. Gyongy. Mimicking the one-dimensional marginal distributions of processes having an Itô differential. *Probab. Th. Rel. Fields*, 71(4):501–516, 1986.

[44] P. Henry-Labordère. Calibration of local stochastic volatility models to market smiles: A Monte-Carlo approach. *Risk Magazine, September*, 2009.

[45] P. Henry-Labordère. (Non)-Parametric Regressions: Applications to Local Stochastic Volatility Models. *Available at SSRN 3374875*, 2019.

[46] C. Jost. A note on ergodic transformations of self-similar Volterra Gaussian processes. *Electron. Commun. Probab.*, 12:259–266, 2007.

[47] B. Jourdain. Loss of martingality in asset price models with lognormal stochastic volatility. *preprint Cermics*, 267:2004, 2004.

[48] R. McCrickerd and M. S. Pakkanen. Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quantitative Finance*, 18(11):1877–1886, 2018.

[49] D. Nualart. *The Malliavin calculus and related topics*, volume 1995. Springer, 2006.

[50] P. Pigato. Extreme at-the-money skew in a local volatility model. *Finance and Stochastics*, 23:827–859, 2019.

[51] M. Romano and N. Touzi. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance*, 7(4):399–412, 1997.

[52] C. A. Sin. Complications with stochastic volatility models. *Advances in Applied Probability*, 30(1):256–268, 1998.

[53] S. Takanobu. Asymptotic expansion formulas of the schilder type for a class of conditional wiener functional integrations. In *Asymptotic problems in probability theory: Wiener functionals and asymptotics, Proceedings of the Taniguchi International Symposium, Sanda and Kyoto, 1990*, pages 194–241. Longman Sci. Tech., 1993.

[54] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Science & Business Media, 2008.