Short dated option pricing under rough volatility

Blanka Horvath

Department of Mathematics, Imperial College London

Advances in Financial Mathematics

Paris, 10-13 Jan 2017

Based on joint works with

P. Harms, A. Jacquier, C. Lacombe and with

C. Bayer, P. Friz, A. Gulisashvili and B. Stemper

Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

Implied volatility

- Asset price process: $(S_t = \mathrm{e}^{X_t})_{t \geq 0}$, with $X_0 = 0$.
- Black-Scholes-Merton (BSM) framework:

$$C_{\mathrm{BS}}(\tau, k, \sigma) := \mathbb{E}_0 \left(\mathrm{e}^{X_\tau} - \mathrm{e}^k \right)_+ = \mathcal{N}(d_+) - \mathrm{e}^k \mathcal{N}(d_-),$$

$$d_\pm := -\frac{k}{\sigma\sqrt{\tau}} \pm \frac{1}{2}\sigma\sqrt{\tau}.$$

- Spot implied volatility $\sigma_\tau(k)$: the unique (non-negative) solution to

$$C_{\mathrm{observed}}(\tau, k) = C_{\mathrm{BS}}(\tau, k, \sigma_\tau(k)).$$

- Implied volatility: unit-free measure of option prices.

Implied volatility

- Asset price process: $(S_t = \mathrm{e}^{X_t})_{t\geq 0}$, with $X_0 = 0$.
- Black-Scholes-Merton (BSM) framework:

$$C_{\mathrm{BS}}(\tau, k, \sigma) := \mathbb{E}_0 \left(\mathrm{e}^{X_\tau} - \mathrm{e}^k\right)_+ = \mathcal{N}(d_+) - \mathrm{e}^k \mathcal{N}(d_-),$$

$$d_\pm := -\frac{k}{\sigma\sqrt{\tau}} \pm \frac{1}{2}\sigma\sqrt{\tau}.$$

- Spot implied volatility $\sigma_\tau(k)$: the unique (non-negative) solution to

$$C_{\mathrm{observed}}(\tau, k) = C_{\mathrm{BS}}(\tau, k, \sigma_\tau(k)).$$

- Implied volatility: unit-free measure of option prices.

Implied volatility is not available in closed form generally.
Its asymptotic behaviour is available via (small/large k, τ) approximations.

Literature

Implied volatility ($\sigma_\tau(k)$) asymptotics as $|k| \uparrow \infty$, $\tau \downarrow 0$ or $\tau \uparrow \infty$:

- Hagan-Kumar-Lesniewski-Woodward (2003/2015), Obłój (2008): small-maturity for the SABR model.
- Berestycki-Busca-Florent (2004): small-τ using PDE methods for diffusions.
- Henry-Labordère (2009): small-τ asymptotics using differential geometry.
- Forde et al.(2012), Jacquier et al.(2012): small/ large τ using large deviations.
- Lee (2003), Benaim-Friz (2009), Gulisashvili (2010-2012), Caravenna-Corbetta (2016), De Marco-Jacquier-Hillairet (2013): $|k| \uparrow \infty$.
- Laurence-Gatheral-Hsu-Ouyang-Wang (2012): small-τ in local volatility models.
- Mijatović-Tankov (2012): small-τ for jump models.
- Bompis-Gobet (2015): asymptotic expansions in the presence of both local and stochastic volatility using Malliavin calculus.
- Fouque et al.(2000-2011): perturbation techniques for slow and fast mean-reverting stochastic volatility models.

Related works:

- Deuschel-Friz-Jacquier-Violante (CPAM 2014), De Marco-Friz (2014): small-noise expansions using Laplace method on Wiener space (Ben Arous-Bismut approach).
- Baudoin-Ouyang (2015): small-noise expansions in a (fully) fractional setting

- Gatheral-Jaisson-Rosenbaum (2014), and Bayer-Gatheral-Friz (2015),
- Forde-Zhang (2015): large deviations in a fractional stochastic volatility setting
- Fukasawa (2011,2015), Alós-León-Vives (2007) small-time (fractional) skew
- Guennoun-Jacquier-Roome (2015), El Euch-Rosenbaum (2016) fractional Heston.

"Classical" case: $(H_i = \frac{1}{2})$ case: Deuschel-Friz-Jacquier-Violante (2011)

$$d\mathbf{X}_t^\varepsilon = b(\varepsilon, \mathbf{X}_t^\varepsilon)dt + \varepsilon \sum_{i=1}^{m} \sigma_i(\mathbf{X}_t^\varepsilon)dW_t^i, \quad \mathbf{X}_0^\varepsilon = \mathbf{x}_0^\varepsilon \in \mathbb{R}^d$$

"Classical" case: $(H_i = \frac{1}{2})$ case: Deuschel-Friz-Jacquier-Violante (2011)

$$d\mathbf{X}_t^\varepsilon = b(\varepsilon, \mathbf{X}_t^\varepsilon)dt + \varepsilon \sum_{i=1}^m \sigma_i(\mathbf{X}_t^\varepsilon) dW_t^i, \quad \mathbf{X}_0^\varepsilon = \mathbf{x}_0^\varepsilon \in \mathbb{R}^d$$

Fractional case: $(H_i = H \in (\frac{1}{4}, 1))$ case: Baudoin-Ouyang (2015)

$$d\mathbf{X}_t^\varepsilon = b(\varepsilon, \mathbf{X}_t^\varepsilon)dt + \varepsilon \sum_{i=1}^m \sigma_i(\mathbf{X}_t^\varepsilon) d(W^H)_t^i, \quad \mathbf{X}_0^\varepsilon = \mathbf{x}_0^\varepsilon \in \mathbb{R}^d$$

"Classical" case: $(H_i = \frac{1}{2})$ case: Deuschel-Friz-Jacquier-Violante (2011)

$$d\mathbf{X}_t^\varepsilon = b(\varepsilon, \mathbf{X}_t^\varepsilon)dt + \varepsilon \sum_{i=1}^m \sigma_i(\mathbf{X}_t^\varepsilon) dW_t^i, \quad \mathbf{X}_0^\varepsilon = \mathbf{x}_0^\varepsilon \in \mathbb{R}^d$$

Fractional case: $(H_i = H \in (\frac{1}{4}, 1))$ case: Baudoin-Ouyang (2015)

$$d\mathbf{X}_t^\varepsilon = b(\varepsilon, \mathbf{X}_t^\varepsilon)dt + \varepsilon \sum_{i=1}^m \sigma_i(\mathbf{X}_t^\varepsilon) d(W^H)_t^i, \quad \mathbf{X}_0^\varepsilon = \mathbf{x}_0^\varepsilon \in \mathbb{R}^d$$

Our main interest: $H_1 = \frac{1}{2}, H_2 \neq \frac{1}{2}$.

Rough volatility models

- Short-term data suggests a time decay of the ATM skew proportional to $\tau^{-\theta}$, with $\theta \in (0, 1/2)$ while classical stochastic volatility models generate a constant short-maturity skew.
- **Gatheral-Jaisson-Rosenbaum** and **Bayer-Gatheral-Friz** (2014,1015) proposed a fractional volatility model:

$$\begin{aligned} \mathrm{d}S_t &= S_t(\sigma_t \mathrm{d}Z_t + \mu_t \mathrm{d}t), \\ \sigma_t &= \exp\left(Y_t\right), \end{aligned} \tag{1}$$

where

$$\mathrm{d}Y_t = \mu \mathrm{d}W_t^H - b(Y_t - m)\mathrm{d}t,$$

for $\mu, b > 0$, $m \in \mathbb{R}$ for a Bm Z and a fBm motion W^H with Hurst parameter H.

- Time series of the Oxford-Man SPX realised variance as well as implied volatility smiles of the SPX suggest that $H \in (0, 1/2)$: **short-memory** volatility.
- Main drawback: loss of Markovianity ($H \neq 1/2$) rules out PDE techniques, and Monte Carlo is computationally intensive. One way out is an efficient "Hybrid scheme" of Bennedsen, Lunde and Pakkanen (2015).

Rough volatility models: our setting

$$dX_t = b_1(Y_t)dt + \sigma_1(Y_t)dW_t$$
$$dY_t = b_2(Y_t)dt + \sigma_2 dW_t^H. \quad (2)$$

$\sigma_2 > 0$ (extendible to bounded, elliptic) and $H \in (0,1)$, particular interest in $H < 1/2$.

Forde-Zhang '16: $b_1 \equiv b_2 \equiv 0$, and $\sigma_1 \in C^\alpha$, $\alpha \in (0,1]$ (second part of the talk);
Fractional Stein-Stein: $b_1(y) \equiv -y^2/2$, $\sigma_1(y) \equiv y$, $b_2(y) \equiv a - by$ (first part).
Gatheral-Jaisson-Rosenbaum '14: $b_1(y) \equiv -e^{2y}/2$, $\sigma_1(y) \equiv e^y$, $b_2(y) \equiv a - by$.

Rough volatility models: our setting

$$dX_t = b_1(Y_t)dt + \sigma_1(Y_t)dW_t$$
$$dY_t = b_2(Y_t)dt + \sigma_2 dW_t^H. \quad (2)$$

$\sigma_2 > 0$ (extendible to bounded, elliptic) and $H \in (0,1)$, particular interest in $H < 1/2$.

Forde-Zhang '16: $b_1 \equiv b_2 \equiv 0$, and $\sigma_1 \in C^\alpha$, $\alpha \in (0,1]$ (second part of the talk);
Fractional Stein-Stein: $b_1(y) \equiv -y^2/2$, $\sigma_1(y) \equiv y$, $b_2(y) \equiv a - by$ (first part).
Gatheral-Jaisson-Rosenbaum '14: $b_1(y) \equiv -e^{2y}/2$, $\sigma_1(y) \equiv e^y$, $b_2(y) \equiv a - by$.

To introduce correlation we consider $\tilde{B}$ and B independent, and set $W = \bar{\rho}\tilde{B} + \rho B$ and

$$W_t^H = \int_0^t K(t,s)dB_s$$

where K the Volterra kernel of the (standard) fBm W^H.

Rough volatility models: our setting

$$dX_t = b_1(Y_t)dt + \sigma_1(Y_t)dW_t$$
$$dY_t = b_2(Y_t)dt + \sigma_2 dW_t^H. \quad (2)$$

$\sigma_2 > 0$ (extendible to bounded, elliptic) and $H \in (0,1)$, particular interest in $H < 1/2$.

Forde-Zhang '16: $b_1 \equiv b_2 \equiv 0$, and $\sigma_1 \in C^\alpha$, $\alpha \in (0,1]$ (second part of the talk);
Fractional Stein-Stein: $b_1(y) \equiv -y^2/2$, $\sigma_1(y) \equiv y$, $b_2(y) \equiv a - by$ (first part).
Gatheral-Jaisson-Rosenbaum '14: $b_1(y) \equiv -e^{2y}/2$, $\sigma_1(y) \equiv e^y$, $b_2(y) \equiv a - by$.

To introduce correlation we consider $\tilde{B}$ and B independent, and set $W = \bar{\rho}\tilde{B} + \rho B$ and

$$W_t^H = \int_0^t K(t,s)dB_s$$

where K the Volterra kernel of the (standard) fBm W^H.

An intuitive remark: Mandelbrot-van Ness representation for fBm:
At time zero, the volatility process in (2) has already accumulated some randomness.
$\Rightarrow$ Hot-start of the process Y (with A. Jacquier and C. Lacombe).

1. Introduction

2. Approach via density asymptotics
    - Varadhan-type asymptotics for fractional SDEs
    - Rescalings and density asymptotics for fractional models
    - Corollaries: Short-time/tail expansion in fractional models
    - Implied volatility asympotics
    - Idea of the proof

3. Bypassing density asymptotics
    - Refined expansions and moderate regimes
    - A non-Markovian extention of Osajima's energy expansion
    - Implied volatility asymptotics

Varadhan-type asymptotics

Recall the Black-Scholes density expansion: heat-kernel asymptotics

$$f_{\mathrm{BS}}(t,x) \sim t^{-1/2} \exp\left(-\frac{1}{2t}\left(\frac{x}{\sigma}\right)^2\right), \quad \text{as } t \to 0.$$

In the **'homogenous" (fractional) case:** $H_1 = H_2 = \ldots = H_d = H$ (Baudoin-Ouyang)

$$d\mathbf{X}_t = b(\mathbf{X}_t)dt + \sum_{i=1}^{m} \sigma_i(\mathbf{X}_t)d(W^H)^i_t, \quad \mathbf{X}_0 = \mathbf{x}_0 \in \mathbb{R}^d \tag{3}$$

(Extended) Varadhan formula

$$f_{\mathrm{X}}(t,x) \sim \text{cst } t^{-H} \exp\left(-\frac{d^2(x_0,x)}{2t^{2H}}\right), \quad \text{as } t \to 0.$$

Varadhan-type asymptotics

Recall the Black-Scholes density expansion: heat-kernel asymptotics

$$f_{\mathrm{BS}}(t,x) \sim t^{-1/2} \exp\left(-\frac{1}{2t}\left(\frac{x}{\sigma}\right)^2\right), \quad \text{as } t \to 0.$$

In the ‘**homogenous" (fractional) case:** $H_1 = H_2 = \ldots = H_d = H$ (Baudoin-Ouyang)

$$d\mathbf{X}_t = b(\mathbf{X}_t)dt + \sum_{i=1}^{m} \sigma_i(\mathbf{X}_t)d(W^H)^i_t, \quad \mathbf{X}_0 = \mathbf{x}_0 \in \mathbb{R}^d \tag{3}$$

(Extended) Varadhan formula

$$f_{\mathrm{X}}(t,x) \sim \mathrm{cst}\ t^{-H} \exp\left(-\frac{d^2(x_0,x)}{2t^{2H}}\right), \quad \text{as } t \to 0.$$

What if Hurst parameters are different $H_1 \neq H_2$?

Rescalings

Key to asymptotic expansions for $H_1 \neq H_2$: Rescalings

Recall the considered processes

$$dX_t = b_1(Y_t)dt + \sigma_1(Y_t)dW_t$$
$$dY_t = b_2(Y_t)dt + \sigma_2 dW_t^H.$$

Rescalings

Key to asymptotic expansions for $H_1 \neq H_2$: Rescalings

Recall the considered processes

$$dX_t = b_1(Y_t)dt + \sigma_1(Y_t)dW_t$$
$$dY_t = b_2(Y_t)dt + \sigma_2 dW_t^H.$$

Define appropriate rescalings

$$dX_t^\varepsilon = b_1(\varepsilon^{\kappa_1}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_1(Y_t^\varepsilon)dW_t$$
$$dY_t^\varepsilon = b_2(\varepsilon^{\kappa_2}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_2 dW_t^H.$$

Rescalings

$$dX_t^\varepsilon = b_1(\varepsilon^{\kappa_1}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_1(Y_t^\varepsilon)dW_t \quad dY_t^\varepsilon = b_2(\varepsilon^{\kappa_2}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_2 dW_t^H. \tag{4}$$

Fractional Stein-Stein: Consider $(X_0, Y_0) = (0, y_0)$ and

$$dX_t = -\frac{Y_t^2}{2}dt + Y_t dW_t, \qquad dY_t = (a - bY_t)dt + c dW_t^H. \tag{5}$$

Rescalings

$$dX_t^\varepsilon = b_1(\varepsilon^{\kappa_1}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_1(Y_t^\varepsilon)dW_t \quad dY_t^\varepsilon = b_2(\varepsilon^{\kappa_2}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_2 dW_t^H. \qquad (4)$$

Fractional Stein-Stein: Consider $(X_0, Y_0) = (0, y_0)$ and

$$dX_t = -\frac{Y_t^2}{2}dt + Y_t dW_t, \qquad dY_t = (a - bY_t)dt + c dW_t^H. \qquad (5)$$

- **Rescaling 1 (short-time):** $(X_t^\varepsilon, Y_t^\varepsilon) := (\varepsilon^{2H-1} X_{\varepsilon^2 t}, Y_{\varepsilon^2 t}) \Rightarrow (4)$ with $\kappa_1 = 2H + 1$, $\kappa_2 = 2$ and $\beta = 2H$, $(x_0^\varepsilon, y_0^\varepsilon) = (0, y_0)$:

$$dX_t^\varepsilon = -\varepsilon^{2H+1}\frac{(Y_t^\varepsilon)^2}{2}dt + \varepsilon^{2H} Y_t^\varepsilon dW_t, \qquad dY_t^\varepsilon = \varepsilon^2(a - bY_t^\varepsilon)dt + \varepsilon^{2H} c dW_t^H. \qquad (6)$$

Rescalings

$$dX_t^\varepsilon = b_1(\varepsilon^{\kappa_1}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_1(Y_t^\varepsilon)dW_t \quad dY_t^\varepsilon = b_2(\varepsilon^{\kappa_2}, Y_t^\varepsilon)dt + \varepsilon^\beta \sigma_2 dW_t^H. \tag{4}$$

Fractional Stein-Stein: Consider $(X_0, Y_0) = (0, y_0)$ and

$$dX_t = -\frac{Y_t^2}{2}dt + Y_t dW_t, \qquad dY_t = (a - bY_t)dt + c dW_t^H. \tag{5}$$

- **Rescaling 1 (short-time):** $(X_t^\varepsilon, Y_t^\varepsilon) := (\varepsilon^{2H-1}X_{\varepsilon^2 t}, Y_{\varepsilon^2 t}) \Rightarrow$ (4) with $\kappa_1 = 2H+1$, $\kappa_2 = 2$ and $\beta = 2H$, $(x_0^\varepsilon, y_0^\varepsilon) = (0, y_0)$:

$$dX_t^\varepsilon = -\varepsilon^{2H+1}\frac{(Y_t^\varepsilon)^2}{2}dt + \varepsilon^{2H} Y_t^\varepsilon dW_t, \qquad dY_t^\varepsilon = \varepsilon^2(a - bY_t^\varepsilon)dt + \varepsilon^{2H} c dW_t^H. \tag{6}$$

- **Rescaling 2 (tails):** $(X_t^\varepsilon, Y_t^\varepsilon) := (\varepsilon^{2H}X_t, \varepsilon^H Y_t) \Rightarrow$ (4) with $\kappa_1 = 0$, $\kappa_2 = \beta = H$, $(x_0^\varepsilon, y_0^\varepsilon) = (0, \varepsilon^H y_0)$:

$$dX_t^\varepsilon = -\frac{(Y_t^\varepsilon)^2}{2}dt + \varepsilon^H Y_t^\varepsilon dW_t, \qquad dY_t^\varepsilon = (a\varepsilon^H + bY_t^\varepsilon)dt + \varepsilon^H c dW_t^H. \tag{7}$$

Theorem (Harms-H-Jacquier)

Consider an SDE of the form (4). Then the density of X_T^ε admits an expansion

$$f_\varepsilon(T,x) = \exp\left(-\frac{\Lambda(x)}{\varepsilon^{2\beta}} + \frac{\Lambda'(x)\widehat{X}_T}{\varepsilon^\beta}\right)\varepsilon^{-\min(\kappa_1,\beta)}\left(c_0 + \mathcal{O}(\varepsilon^{\delta(\kappa_1,\beta)})\right), \quad \text{as } \varepsilon \to 0,$$

where

$$\Lambda(x) = \inf\left\{\frac{1}{2}\|\mathrm{k}\|^2_{\mathcal{H}_H}, \mathrm{k} \in \mathcal{K}^x_{x_0^0, y_0^0}\right\} = \frac{1}{2}\|\mathrm{k}_0\|^2_{\mathcal{H}_H},$$

and

$$\mathrm{d}\widehat{X}_t = \left[\partial_x b_1\left(0, \phi_t^{\mathrm{k}_0}\right) + \partial_x \sigma_1\left(\phi_t^{\mathrm{h}_0}\right)\cdot \dot{\mathrm{k}}_0(t)\right]\widehat{X}_t \mathrm{d}t + \partial_{\varepsilon^\beta} b_1\left(0, \phi_t^{\mathrm{k}_0}\right)\mathrm{d}t, \; \widehat{X}_0 = \partial_{\varepsilon^\beta} x_0^\varepsilon|_{\varepsilon=0},$$

where ϕ^{k_0} denotes the ODE solution of the same SDE (4) replacing $\varepsilon^\beta \mathrm{d}W$ by $\dot{\mathrm{k}}_0$ and x_0^ε by x_0^0.

Notations

- $\mathcal{H}$: absolutely continuous paths $[0,T] \to \mathbb{R}^2$ starting at 0 such that $\left\|\dot{\mathrm{h}}\right\|_{\mathcal{H}}^2 < \infty$.
- $\mathcal{H}_H := K_H \mathcal{H}$ and $\mathrm{k} := K_H \mathrm{h}$, where K_H denotes the Volterra kernel.
- For fixed $(x_0, y_0) \in \mathbb{R}^2$, ϕ^{k} is the (unique) ODE solution to

$$\dot{\phi}_t^{\mathrm{k}} = b_1\left(0, \phi_t^{\mathrm{k}}\right) \mathrm{d}t + \sigma_1\left(\phi_t^{\mathrm{h}}\right) \mathrm{dk}_t^1 + \sigma_2\left(\phi_t^{\mathrm{h}}\right) \mathrm{dk}_t^2, \quad \phi_0^{\mathrm{k}} = (x_0^0, y_0^0).$$

- Denote $\psi^{\mathrm{k}} := \Pi_1 \phi^{\mathrm{k}}$ its projection on to the first coordinate X.
- $\mathcal{K}_{\mathrm{a}} := \left\{\mathrm{k} \in \mathcal{H}_H : \psi_T^{\mathrm{k}} = a \in \mathbb{R}\right\} \neq \emptyset$ ("by Hörmander condition").
- $\Lambda(\mathrm{a}) := \inf\left\{\frac{1}{2}\|\mathrm{k}\|_H^2 : \mathrm{k} \in \mathcal{K}^{\mathrm{a}}\right\}$.

Corollary: Varadhan-type asymptotics

Corollary (short-time asymptotics in Stein-Stein) $dY_t = (a - bY_t)dt + cdW_t^H$

In the fractional Stein-Stein model (X_t, Y_t) with $X_0 = 0$, $Y_0 = y_0 > 0$ the density of X_t satisfies in a neighbourhood of (x_0, y_0) the following asymptotic expansion as $t \to 0$

$$f_X(t,x) = \exp\left(-\frac{\Lambda(x)}{t^{2H}}\right) t^{-H}\left(\frac{1}{2\pi} + \mathcal{O}(t^{\delta(H,H+1/2)})\right)$$

where

$$\Lambda(x) = \inf\left\{\frac{1}{2}\|\mathrm{k}\|^2_{\mathcal{H}_H}, \mathrm{k} \in \mathcal{K}^x_{x_0,y_0}\right\}.$$

Corollary: Varadhan-type asymptotics

Corollary (short-time asymptotics in Stein-Stein) $dY_t = (a - bY_t)dt + cdW_t^H$

In the fractional Stein-Stein model (X_t, Y_t) with $X_0 = 0$, $Y_0 = y_0 > 0$ the density of X_t satisfies in a neighbourhood of (x_0, y_0) the following asymptotic expansion as $t \to 0$

$$f_X(t,x) = \exp\left(-\frac{\Lambda(x)}{t^{2H}}\right) t^{-H} \left(\frac{1}{2\pi} + \mathcal{O}(t^{\delta(H,H+1/2)})\right)$$

where

$$\Lambda(x) = \inf\left\{\frac{1}{2}\, \|\mathrm{k}\|^2_{\mathcal{H}_H}, \mathrm{k} \in \mathcal{K}^x_{x_0,y_0}\right\}.$$

Proof: Take $T = 1$, $\varepsilon^2 = t$ and consider $(X_t^\varepsilon, Y_t^\varepsilon) := (\varepsilon^{2H-1} X_{\varepsilon^2 t}, Y_{\varepsilon^2 t})$ with $X_0^\varepsilon = 0$, $Y_0^\varepsilon = y_0 > 0$. $\Rightarrow$ Short-time scaling:

$$dX_t^\varepsilon = -\varepsilon^{2H+1}\frac{(Y_t^\varepsilon)^2}{2}dt + \varepsilon^{2H} Y_t^\varepsilon dW_t, \qquad dY_t^\varepsilon = \varepsilon^2(a + bY_t^\varepsilon)dt + \varepsilon^{2H} cdW_t^H, \quad (6)$$

Note that the drift vanishes in the limit $\varepsilon \to 0$ and $x_0^\varepsilon = x_0 = 0$.
$\Rightarrow$ $(\widehat{X}_t, \widehat{Y}_t) \equiv 0$, so that there is no $1/\varepsilon^\beta = 1/t^{\beta/2}$ term in the exponential.

Corollary: tail asymptotics

Corollary (tail expansion in Stein-Stein) $dY_t = (a + bY_t)dt + cdW_t^H$

Consider the fractional Stein-Stein model with $X_0 = 0$, $Y_0 = y_0 > 0$. Then as $x \to \infty$,

$$f_X(T, x) = \exp\left(-c_1 x + c_2 x^{1/2}\right) \frac{1}{x^{1/2}} \left(c_0 + \mathcal{O}\left(x^{1/2}\right)\right)$$

where $c_1 := \Lambda(1)$, $c_2 := \widehat{X}_T \Lambda'(1)$.

Note that the expression on the RHS is independent of the Hurst-parameter!

Corollary: tail asymptotics

Corollary (tail expansion in Stein-Stein) $dY_t = (a + bY_t)dt + cdW_t^H$

Consider the fractional Stein-Stein model with $X_0 = 0$, $Y_0 = y_0 > 0$. Then as $x \to \infty$,

$$f_X(T,x) = \exp\left(-c_1 x + c_2 x^{1/2}\right) \frac{1}{x^{1/2}} \left(c_0 + \mathcal{O}\left(x^{1/2}\right)\right)$$

where $c_1 := \Lambda(1)$, $c_2 := \widehat{X}_T \Lambda'(1)$.

Note that the expression on the RHS is independent of the Hurst-parameter!

Proof: Consider $(X_T^\varepsilon, Y_T^\varepsilon) := (\varepsilon^{2H} X_T, \varepsilon^H Y_T)$ with $X_0^\varepsilon = \varepsilon^{2H} X_0$ and $Y_0^\varepsilon = \varepsilon^H Y_0$.

$$dX_t^\varepsilon = -\frac{(Y_t^\varepsilon)^2}{2} dt + \varepsilon^H Y_t^\varepsilon dW_t, \qquad dY_t^\varepsilon = (a\varepsilon^H + bY_t^\varepsilon)dt + \varepsilon^H c dW_t^H, \qquad (7)$$

Note that $X_T^\varepsilon \stackrel{\Delta}{=} \varepsilon^{2H} X_T$. $\Rightarrow \mathbb{P}(X_T^\varepsilon \geq y) = \mathbb{P}(X_T \geq y/\varepsilon^{2H})$, $\Rightarrow$
$f_X(T, y/\varepsilon^{2H}) = \varepsilon^{2H} f_\varepsilon(T,y)$. Take $y = 1$, that is $x := \varepsilon^{-2H}$. By the theorem,
$f_\varepsilon(T,1) \approx \exp\left(-\frac{\Lambda(1)}{\varepsilon^{2H}} + \dots\right) \frac{1}{\varepsilon^H}$, hence
$f_X(T,x) \approx \exp\left(-\frac{\Lambda(1)}{\varepsilon^{2H}} + \dots\right) \varepsilon^H = \exp\left(-\Lambda(1)x + \dots\right) \frac{1}{x^{1/2}}$.

From density to implied volatility: small-time

Recall the Black-Scholes density expansion:

$$f_{\mathrm{BS}}(t,x) \sim t^{-1/2} \exp\left(-\frac{1}{2t}\left(\frac{x}{\sigma}\right)^2\right), \quad \text{as } t \to 0, \text{ for any } x \in \mathbb{R}.$$

The corollary (Varadhan-type asymptotics) implies that in the fractional Stein-Stein model

$$f_{\mathrm{X}}(t,x) \sim \text{cst } t^{-H} \exp\left(-\frac{d^2(x_0, y_0; x)}{2t^{2H}}\right), \quad \text{as } t \to 0.$$

From density to implied volatility: small-time

Recall the Black-Scholes density expansion:

$$f_{\mathrm{BS}}(t,x) \sim t^{-1/2} \exp\left(-\frac{1}{2t}\left(\frac{x}{\sigma}\right)^2\right), \quad \text{as } t \to 0, \text{ for any } x \in \mathbb{R}.$$

The corollary (Varadhan-type asymptotics) implies that in the fractional Stein-Stein model

$$f_{\mathrm{X}}(t,x) \sim \text{cst } t^{-H} \exp\left(-\frac{d^2(x_0,y_0;x)}{2t^{2H}}\right), \quad \text{as } t \to 0.$$

Matching the leading-orders gives

$$\sigma_{\mathrm{BS}}(t,x) \sim \frac{|x|}{d(x_0,y_0;x)} t^{H-1/2} \quad \text{as } t \to 0.$$

Skew explodes with rate $H - 1/2$ in the short end whenever $H < 1/2$.

From density to implied volatility: tails

Recall the Black-Scholes density expansion:

$$f_{\mathrm{BS}}(t,x) \sim \exp\left(-\frac{x^2}{2\sigma^2 t} - \frac{x}{4}\right) \quad \text{as } x \to \infty, \text{ for any } t > 0.$$

Our theorem (corollary) says that in the fractional Stein-Stein model (5), we have

$$f_{\mathrm{X}}(t,x) \sim \frac{\mathrm{cst}}{x^{1/2}} \exp\left(-c_1 x + c_2\sqrt{x}\right), \quad \text{as } x \to \infty.$$

From density to implied volatility: tails

Recall the Black-Scholes density expansion:

$$f_{\mathrm{BS}}(t,x) \sim \exp\left(-\frac{x^2}{2\sigma^2 t} - \frac{x}{4}\right) \quad \text{as } x \to \infty, \text{ for any } t > 0.$$

Our theorem (corollary) says that in the fractional Stein-Stein model (5), we have

$$f_{\mathrm{X}}(t,x) \sim \frac{\text{cst}}{x^{1/2}} \exp\left(-c_1 x + c_2\sqrt{x}\right), \quad \text{as } x \to \infty.$$

Matching the leading-orders gives

$$-c_1 x + c_2\sqrt{x} \sim -\frac{x^2}{2\sigma^2 t} - \frac{x}{4},$$

and we recover Roger Lee's formula independently of the Hurst exponent in (5).

Proof of Theorem 1

$$dX_t = -\varepsilon^{2H+1}\frac{1}{2}Y_t^2 dt + \varepsilon^{2H} Y_t dW_t, \qquad dY_t = \varepsilon^{2H} dW_t^H,$$

with the same initial condition $X_0 = Y_0 = 0$.

$$\text{Density: } f_\varepsilon(T,x) = \exp\left[-\frac{\Lambda(x)}{\varepsilon^{4H}} + \frac{\Lambda'(x)\widehat{X}_T}{\varepsilon^{2H}}\right]\varepsilon^{-2H}\left(c_0 + \mathcal{O}(\varepsilon^{2H})\right).$$

Proof of Theorem 1

$$dX_t = -\varepsilon^{2H+1}\frac{1}{2}Y_t^2 dt + \varepsilon^{2H} Y_t dW_t, \qquad dY_t = \varepsilon^{2H} dW_t^H,$$

with the same initial condition $X_0 = Y_0 = 0$.

$$\text{Density: } f_\varepsilon(T,x) = \exp\left[-\frac{\Lambda(x)}{\varepsilon^{4H}} + \frac{\Lambda'(x)\widehat{X}_T}{\varepsilon^{2H}}\right]\varepsilon^{-2H}\left(c_0 + \mathcal{O}(\varepsilon^{2H})\right).$$

Proof: Take $x \in \mathbb{R}$ and a C^∞-bounded function F such that $F(x) = 0$.

$$f_\varepsilon(T,x)\mathrm{e}^{-F(x)/\varepsilon^{4H}} = \frac{1}{2\pi\varepsilon^{2H}}\int_{\mathbb{R}}\mathbb{E}\left\{\exp\left[\mathrm{i}(\zeta,0)\cdot\left(\frac{X_T^\varepsilon-(x,0)}{\varepsilon^{2H}}\right) - \frac{F(X_T^\varepsilon)}{\varepsilon^{4H}}\right]\right\}\mathrm{d}\zeta.$$

Choose F such that $F(\cdot) + \Lambda_{x_0}(\cdot)$ has a non-degenerate minimum at z. This implies that $\mathrm{k} \mapsto F(\phi_T^{\mathrm{k}}(x_0,y_0)) + \frac{1}{2}\|\mathrm{k}\|^2_{\mathcal{H}_H}$ has a non-degenerate minimum at $\mathrm{k}_0 \in \mathcal{H}_H$.

(For instance $F(z) = \lambda|z-x|^2 - [\Lambda_{x_0,y_0}(z) - \Lambda_{x_0,y_0}(x)]$ with $\lambda > 0$).

Proof of Theorem 1

Replace $\varepsilon^{2H}\mathrm{d}B$ ($B := (W, W^H)$) in the SDE by $\varepsilon^{2H}\mathrm{d}W + \dot{\mathrm{k}}_0$.
Call the corresponding Girsanov-transformed process $\widetilde{Z}_t^\varepsilon = (\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon)$:

$$d\widetilde{X}_t^\varepsilon = -\varepsilon^{2H+1}\frac{1}{2}\widetilde{Y}_t^2 dt + \widetilde{Y}_t^\varepsilon(\varepsilon^{2H}dW_t + (\dot{\mathrm{k}}_0)_1), \qquad d\widetilde{Y}_t = \varepsilon^{2H}dW_t^H + (\dot{\mathrm{k}}_0)_2.$$

Girsanov factor

$$\mathcal{G} = \exp\left(-\frac{1}{\varepsilon^{2H}}\int_0^T \psi(\mathrm{k}_0)_t dB_t - \frac{1}{2\varepsilon^{4H}}\|\mathrm{k}_0\|^2_{\mathcal{H}_H}\right).$$

Therefore

$$f(x,T)e^{-F(x)/4\varepsilon^{4H}} = \frac{1}{2\pi\varepsilon^{2H}}\int_{\mathbb{R}}\mathbb{E}\left[e^{\varepsilon^{2H}i\zeta(\widetilde{X}_T - x) - \varepsilon^{-4H}F(\widetilde{X}_T)}\mathcal{G}\right]d\zeta$$
$$= \frac{1}{2\pi\varepsilon^{2H}}\int_{\mathbb{R}}\mathbb{E}\left[e^{(*)}\right]d\zeta$$

where

$$(*) = \varepsilon^{2H}i\zeta(\widetilde{X}_T - x) - \varepsilon^{-4H}F(\widetilde{X}_T) - \varepsilon^{-2H}\int_0^T \psi(\gamma)_t dB_t - \varepsilon^{-4H}\frac{1}{2}\|\gamma\|^2_{\mathcal{H}_H}.$$

Proof of Theorem 1

Replace $\varepsilon^{2H}\mathrm{d}B$ ($B := (W, W^H)$) in the SDE by $\varepsilon^{2H}\mathrm{d}W + \dot{\mathrm{k}}_0$.
Call the corresponding Girsanov-transformed process $\widetilde{Z}_t^\varepsilon = (\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon)$:

$$d\widetilde{X}_t^\varepsilon = -\varepsilon^{2H+1}\frac{1}{2}\widetilde{Y}_t^2 dt + \widetilde{Y}_t^\varepsilon(\varepsilon^{2H}dW_t + (\dot{\mathrm{k}}_0)_1), \qquad d\widetilde{Y}_t = \varepsilon^{2H}dW_t^H + (\dot{\mathrm{k}}_0)_2.$$

Girsanov factor

$$\mathcal{G} = \exp\left(-\frac{1}{\varepsilon^{2H}}\int_0^T \psi(\mathrm{k}_0)_t dB_t - \frac{1}{2\varepsilon^{4H}}\|\mathrm{k}_0\|^2_{\mathcal{H}_H}\right).$$

By a stochastic Taylor expansion of $\widetilde{Z}_t^\varepsilon = (\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon)$ for $\varepsilon^{2H} \to 0$,

$$\exp\left(\frac{-F\left(\widetilde{X}_t^\varepsilon\right)}{\varepsilon^{4H}}\right) = \exp\left[\frac{-1}{\varepsilon^{4H}}\left(F(x) - \varepsilon^{2H}\int_0^T \psi(\mathrm{k}_0)_t \mathrm{d}B_t - \varepsilon^{2H}\widehat{X}_T \cdot \Lambda'_{x_0}(x) + \mathcal{O}(\varepsilon^{4H})\right)\right]$$

Proof of Theorem 1

Replace $\varepsilon^{2H}\mathrm{d}B$ ($B := (W, W^H)$) in the SDE by $\varepsilon^{2H}\mathrm{d}W + \dot{\mathrm{k}}_0$.
Call the corresponding Girsanov-transformed process $\widetilde{Z}_t^\varepsilon = (\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon)$:

$$d\widetilde{X}_t^\varepsilon = -\varepsilon^{2H+1}\frac{1}{2}\widetilde{Y}_t^2 dt + \widetilde{Y}_t^\varepsilon(\varepsilon^{2H}dW_t + (\dot{\mathrm{k}}_0)_1), \qquad d\widetilde{Y}_t = \varepsilon^{2H}dW_t^H + (\dot{\mathrm{k}}_0)_2.$$

Girsanov factor

$$\mathcal{G} = \exp\left(-\frac{1}{\varepsilon^{2H}}\int_0^T \psi(\mathrm{k}_0)_t dB_t - \frac{1}{2\varepsilon^{4H}}\|\mathrm{k}_0\|^2_{\mathcal{H}_H}\right).$$

By a stochastic Taylor expansion of $\widetilde{Z}_t^\varepsilon = (\widetilde{X}_t^\varepsilon, \widetilde{Y}_t^\varepsilon)$ for $\varepsilon^{2H} \to 0$,

$$\exp\left(\frac{-F\left(\widetilde{X}_t^\varepsilon\right)}{\varepsilon^{4H}}\right) = \exp\left[\left(\frac{1}{\varepsilon^{2H}}\int_0^T \psi(\mathrm{k}_0)_t dB_t + \frac{1}{\varepsilon^{2H}}\widehat{X}_T \cdot \Lambda'_{x_0}(x) + \mathcal{O}(1)\right)\right].$$

The rest of the proof follows Ben Arous' proof for X_T^ε.

1 Introduction

2 Approach via density asymptotics
- Varadhan-type asymptotics for fractional SDEs
- Rescalings and density asymptotics for fractional models
- Corollaries: Short-time/tail expansion in fractional models
- Implied volatility asympotics
- Idea of the proof

3 Bypassing density asymptotics
- Refined expansions and moderate regimes
- A non-Markovian extention of Osajima's energy expansion
- Implied volatility asymptotics

Bypassing density asymptotics

Direct call price expansion: For $x \geq 0$ the option price satisfies

$$c(\varepsilon^{1-2H}x, t) := E\left[(\exp(X_t) - \exp(\varepsilon^{1-2H}x))^+\right]$$
$$= E\left[(\exp(Z_t) - \exp(\varepsilon^{1-2H}x))^+ \; G\big|_*\right]$$

where Z is the controlled process around the optimal path k and $G\big|_* = e^{\frac{-I(x)}{\varepsilon^{4H}}} e^{\frac{-I'(x)g_1}{\varepsilon^{2H}}}$ is the Girsanov factor for the optimal path, and g_1 a Gaussian random variable. Then

$$c(\varepsilon^{1-2H}x, t) = \exp\left(-\frac{I(x)}{\varepsilon^{4H}}\right)\exp\left(x\varepsilon^{1-2H}\right)J(\varepsilon, x), \quad \text{where} \quad \widehat{U} := \widehat{Z}_1^{\varepsilon} - x \text{ and}$$

$$J(\varepsilon, x) := E\left[\exp\left(\frac{-I'(x)}{\varepsilon^{2H}}\;\widehat{U}\right)\left(\exp(\varepsilon^{1-2H}\widehat{U}) - 1\right)\exp\left(I'(x)R_2\right)\mathbf{1}_{\widehat{U}\geq 0}\right].$$

$\Rightarrow$ Expansion (uniformly in x) via epansion of the “energy” I directly (Osajima).

Moderate regimes

- Moderate Regimes (in the sense of Friz-Gerhold-Pinter '16) **interpolate** between **out-of-the-money** calls with fixed strike $\left(\log \frac{K}{S_0}\right) = k > 0$ and **at-the-money** $k = 0$ calls: Now $k_t = ct^\theta \Rightarrow$ MOTM (for $0 < \theta < \frac{1}{2}$) and AATM (for larger θ)
- **Reflects market reality**: options closer expiry $\Rightarrow$ strikes closer to the money first observed by Mijatović-Tankov on FX markets
- The moderate regime (MOTM) **permits explicit computations** for the rate function $\Lambda(k)$ in terms of the model parameters
  Moderate deviations $\Rightarrow$ **Advantage over large deviations** (OTM) case where the $\Lambda(k)$ often related to geodesic distance problems and not explicitly available.
- MOTM expansions naturally involve **quantities very familiar to practitioners**, notably spot (implied) volatility, implied volatility skew . . .
- In some cases (fractional volatility models) the scaling θ permits a **fine-tuning** to understand the behavior and derivatives of the energy function.

Moderate regimes for rough volatility

Rescalings $\Longrightarrow$ We tacitly agreed to consider $\mathbb{P}\left(X_t \approx t^{1/2-H}x\right)$. Now it is only a small step to consider instead (for some suitable small $\theta > 0$)

$$\mathbb{P}\left(X_t \approx t^{1/2-H+\theta}x\right).$$

Moderate regimes for rough volatility

Rescalings $\implies$ We tacitly agreed to consider $\mathbb{P}\left(X_t \approx t^{1/2-H}x\right)$. Now it is only a small step to consider instead (for some suitable small $\theta > 0$)

$$\mathbb{P}\left(X_t \approx t^{1/2-H+\theta}x\right).$$

Theorem (Bayer-Friz-Gulisashvili-H-Stemper)

Consider a moderately out-of-the-money call $k_t = xt^{1/2-H+\theta}$; $\theta \in (0, H)$ resp. $\theta \in (0, \frac{2H}{3})$. Then as $t \to 0$, the following holds

$$\log c(k_t, t) \approx \frac{1}{2}\Lambda''(0)\frac{x^2}{t^{2H-2\theta}} + \frac{1}{6}\Lambda'''(0)\frac{x^3}{t^{2H-3\theta}},$$

where we have explicit expressions: $\Lambda''(0) = \frac{1}{\sigma_0^2}$ and $\Lambda'''(0) = -\rho\frac{6\sigma_0'}{\sigma_0^4}\langle K, 1\rangle$.

Here K denotes the Volterra kernel and $\langle K, 1\rangle := \int_0^1 \int_0^t K(t,s)dsdt$.

Moderate regimes for rough volatility

Corollary (MOTM Implied volatility skew)

In the moderately out-of-the-money case ($k_t = xt^{1/2-H+\theta}$; $\theta \in (0, H)$; $\theta \in (0, \frac{2H}{3})$) the implied volatility satisfies the expansion

$$\sigma_{impl}(k_t, t) = \sigma_0 - \rho \frac{\sigma_0'}{\sigma_0} \int_0^1 \int_0^t K(t,s)\ dsdt\ k_t\ t^{H-1/2}\Big(1 + o(1)\Big).$$

Moderate regimes for rough volatility

Corollary (MOTM Implied volatility skew)

In the moderately out-of-the-money case ($k_t = xt^{1/2-H+\theta}$; $\theta \in (0, H)$; $\theta \in (0, \frac{2H}{3})$) the implied volatility satisfies the expansion

$$\sigma_{impl}(k_t, t) = \sigma_0 - \rho \frac{\sigma_0'}{\sigma_0} \int_0^1 \int_0^t K(t,s)\ dsdt\ k_t\ t^{H-1/2}\Big(1 + o(1)\Big).$$

Proof: The statement follows from the Theorem via matching components in the asymptotic expansions and by (see Gao-Lee (2014)) using

$$t\sigma^2_{impl}(k_t, t) \approx \frac{-k_t^2}{2 \log c(k_t, t)}.$$

Moderate regimes for rough volatility

Corollary (MOTM Implied volatility skew)

In the moderately out-of-the-money case ($k_t = xt^{1/2-H+\theta}$; $\theta \in (0, H)$; $\theta \in (0, \frac{2H}{3})$) the implied volatility satisfies the expansion

$$\sigma_{impl}(k_t, t) = \sigma_0 - \rho \frac{\sigma_0'}{\sigma_0} \int_0^1 \int_0^t K(t,s)\ dsdt\ k_t\ t^{H-1/2}\Big(1 + o(1)\Big).$$

Proof: The statement follows from the Theorem via matching components in the asymptotic expansions and by (see Gao-Lee (2014)) using

$$t\sigma^2_{impl}(k_t, t) \approx \frac{-k_t^2}{2 \log c(k_t, t)}.$$

This formula for the skew is in accord with ones previously derived by Alòs-León-Vives (2007) and Fukasawa (2011, 2016) in different settings.

Thank you!