Small-time VIX smile and the stationary distribution for the Rough Heston model

Martin Forde Stefan Gerhold* Benjamin Smith†

December 13, 2022

Abstract

We characterize the asymptotic behaviour of close-to-the-money VIX call options under a generalized Rough Heston model with an exogenously specified initial variance term structure $\xi_0(.)$ and strike $\mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}}$ as the maturity $T \to 0$, using an explicit formula for sampling VIX_T. Using the usual definition $\mathrm{VIX}_T^2 := \frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(u)du$ where $\xi_t(u) := \mathbb{E}(V_u|\mathcal{F}_t)$, we find that $(\mathrm{VIX}_T^2 - \mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}$ satisfies a large deviation principle (LDP) as $T \to 0$ with rate function $I^{\rho=1}(\frac{x}{c})$ and speed T^{-2H} where $c = \frac{\nu}{\Gamma(\alpha)(\frac{1}{2}+H)}\Delta^{H-\frac{1}{2}}$ and $I^{\rho=1}(.)$ is the corresponding rate function for the re-scaled log stock price $(X_T - X_0)/T^{\frac{1}{2}-H}$ in the main Theorem 3.3 in [FGS21] for the special case where $\rho = 1$. This implies that the VIX smile exhibits maximal positive power-law skew as $T \to 0$ in some sense, and we compute an explicit small log-moneyess expansion for the asymptotic smile which shows that asymptotic VIX implied volatility has positive skew and negative convexity at-the-money, consistent with empirical observations. In the final section, we also formally show that V_t has a well defined asymptotic distribution as $t \to \infty$, and we give an explicit formula for the mgf of V_∞ in terms of the solution to a VIE.[1]

1 Introduction

The Rough Heston stochastic volatility model was introduced in Jaisson&Rosenbaum[JR16], and (using C-tightness arguments from Jacod&Shiryaev[JS13]) they show that the model arises naturally as a weak large-time limit of a high-frequency market microstructure model driven by two nearly unstable Hawkes process. [ER19] show that the characteristic function of the log stock price for the Rough Heston model admits a quasi-closed form solution via the solution to a non-linear Volterra integral equation (VIE) (see also [EFR18] and [ER18]), and the variance curve for the model evolves as $d\xi_u(t) = \kappa(u-t)\sqrt{V_t}dW_t$, where $\kappa(t)$ is the usual fractional kernel $t^{H-\frac{1}{2}}$ for the V process multiplied by a *Mittag-Leffler* function. The instantaneous variance process V for the model is $(H-\varepsilon)$-Hölder continuous like fractional Brownian motion (see e.g. Theorem 3.2 in [JR16]) and the model exhibits power law skew in the small-time limit (see Theorem 3.1 in [FGS21] and Corollary 3.4 in [FSV21]). [DJR19] introduce an extension of this model known as the *super Rough Heston model* which incorporates the empirically observed *strong Zumbach effect* as a weak limit of a market microstructure model driven by a quadratic Hawkes process (also using C-tightness arguments) but this model is no longer affine and thus not directly amenable to VIE techniques or Edgeworth and large deviation asymptotics, so it is difficult to prove anything about the qualitative behaviour or dynamics of the smile (and the Zumbach term is a drift term and hence very unlikely to affect leading order large deviation asymptotics). A variant of this model is used in [GJR20], which attains a better fit to SPX and VIX options in practice than conventional rough volatility models, but Guyon[Guy20b] remarks if we calibrate this model to the VIX smile, the short-maturity at-the-money SPX skew is still too small compared to what is observed in practice (see below for discussion on the addition of jumps in [FS21]).

The theoretical value of the VIX index at time t is $\mathrm{VIX}_t = \sqrt{-\frac{2}{\Delta}\mathbb{E}(\log\frac{S_{t+\Delta}}{S_t}|\mathcal{F}_t)}$, where S_t is the S&P 500 index value at time t, $\Delta = 30$ days and $\mathcal{F}_t$ is the market filtration, so VIX_t^2 is effectively a rolling 30-day Variance swap rate. A VIX option is a European call or put option on VIX_T for some maturity T, and if we replace the spot value S_0 in the Black-Scholes formula with the VIX future price $\mathbb{E}(\mathrm{VIX}_T)$, we can define the implied volatility of a VIX call or put in the usual way by inverting the Black-Scholes formula. VIX options are very liquid in practice (although their bid/offer spreads

*TU Wien, Financial and Actuarial Mathematics, Wiedner Hauptstraße 8/105-1, A-1040 Vienna, Austria (sgerhold@fam.tuwien.ac.at)

†Dept. Mathematics, King's College London, Strand, London, WC2R 2LS (Benjamin.Smith@kcl.ac.uk)

[1] We would like to thank Eduardo Abi Jaber, Florian Bourgey, Julien Guyon and Alexandre Pannier for explaining many relevant and useful results in their articles

are still comparatively high), and empirical VIX smile typically exhibit positive skew with negative convexity (see plots in [GJR20],[Guy20],[DeM18],[HJT20] et al.), although e.g. Markovian diffusion models like the standard Heston model can give rise to negative VIX implied vol skews..

In this article, we work with a generalized version of the Rough Heston model as used in [GR19] with initial variance curve $\xi_0(t)$ and the corresponding dynamics of the forward variance $\xi_t(u)$. We first derive an explicit formula for simulating VIX_T in Eq (3) (note that no such formula exists for e.g. the quadratic rough Heston model in [GJR20]) and we then perform a formal small T-expansion of $\xi_T(u)$ which suggests that $(\mathrm{VIX}_T^2-\mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}\sim c\tilde{X}_T/T^{\frac{1}{2}-H}$ in some sense as $T\to 0$ for some constant $c>0$, where $\tilde{X}_t=\int_0^t\sqrt{V_s}dW_s$ is the martingale component of the log stock price for the driftless rough Heston model when the correlation $\rho=1$. This leads us to guess that $(\mathrm{VIX}_T^2-\mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}$ satisfies the same small-time LDP as $c\tilde{X}_T/T^{\frac{1}{2}-H}$, for which we can readily compute a small-time LDP with minor amendments to the main arguments in [FGS21] to allow for non-flat $\xi_0(t)$. We then make this rigorous by showing that $(\mathrm{VIX}_T^2-\mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}$ and $c\tilde{X}_T/T^{\frac{1}{2}-H}$ are *exponentially equivalent* as $T\to 0$ and hence satisfy the same LDP, and this is proved using a minor variant/extension of Theorem 7.1 in Abi Jaber et al.[ALP19] for the exponential-affine formula for $\mathbb{E}(e^{u\tilde{X}_T+(f*\tilde{X})_T}|\mathcal{F}_t^W)$ for a general function f and $u\in\mathbb{R}$ such that T is less than the explosion time $T^*(u)$. Specifically we show that $\varepsilon^{2H}\log\mathbb{E}(e^{\varepsilon^{-\alpha}p(\mathrm{VIX}_\varepsilon^2-\mathrm{VIX}_0^2-c\tilde{X}_\varepsilon)})=V_0I^{1-\alpha}\phi_\varepsilon(p,1)$ where $\alpha=H+\frac{1}{2}$, and $\phi_\varepsilon(p,t)$ satisfies a family of VIEs whose solution tends uniformly to zero on $[0,1]$ as $\varepsilon\to 0$ for all $p\in\mathbb{R}$ and I^r denotes the r-th order fractional integral operator. We later translate this LDP into VIX call option and implied volatility asymptotics, and we compute a small log-moneyness expansion for the asymptotic VIX smile using expansions previously derived in [FGS21] which yields tractable expressions for the overall level, skew and convexity of the short-end VIX smile. We also mention Proposition 18 in [AGM18] which shows that the derivative of the VIX implied volatility with respect to log-moneyness at-the-money tends to a finite constant as $T\to 0$ (as opposed to exploding power-law behaviour $\propto T^{H-\frac{1}{2}}$) for a standard (and mixed) rough Bergomi-type model, and we have verified this behaviour numerically. However numerical computations suggest that for the rough Heston model, the [AGM18] measure of at-the-money skew does indeed appear to be $O(T^{H-\frac{1}{2}})$ as $T\to 0$, as one would guess from (19) below.

Unfortunately, since the limiting VIX smile only depends on the factor $\nu/\sqrt{V_0}$ and not on ρ, we cannot simultaneously fit the overall level and skew of observed limiting VIX smile using the standard rough Heston model. To circumvent this issue, the companion article [FS21] enriches the model with an additional independent CGMY (a.k.a. KoBoL)-type Lévy process L as in [FSV21] with $Y\in(1,2)$, and using a simple modification of the main result in [FSV21] for the Edgeworth regime where log-moneyness scales like $x\sqrt{T}$, we show that one can simultaneously use the rough Heston parameters to fit the at-the-money VIX level and skew as $T\to 0$, and the CGMY parameters to fit the observed level, at-the-money correction and at-the-money skew of SPX options as $T\to 0$ (using the main Theorem in [FSV21] adapted for our rough Heston V process), and the drift of the V process can be made to be fully consistent with the initial observed variance curve structure.

In section 3, we formally show that V_t has a well defined stationary distribution as $t\to\infty$, and we give a semi-explicit formula for its mgf in terms of the solution to a non-standard non-linear VIE, and we verify that the result is consistent with classical result that V_∞ has a Gamma distribution when $H=\frac{1}{2}$. We also compute the law of V_∞ explicitly for a rough-Bergomi type model with a Gamma kernel.

2 The model

We consider a generalized Rough Heston model for a log stock price process $X_t=\log S_t$ of the same form in Gatheral&Radoičič[GR19]:

$$\begin{aligned} dX_t &= -\frac{1}{2}V_t dt+\sqrt{V_t}(\rho dW_t+\bar{\rho}dB_t) \\ V_t &= \xi_0(t)\,+\,c_\alpha\int_0^t(t-s)^{\alpha-1}\nu\sqrt{V_s}dW_s \end{aligned} \tag{1}$$

for $H\in(0,\frac{1}{2})$, $\alpha=H+\frac{1}{2}$, $c_\alpha=\frac{1}{\Gamma(\alpha)}$ and $\nu>0$, with some initial variance curve $\xi_0(t)$ with $\xi_0(.)$ continuous, where W, B are two independent Brownian motions, $\bar{\rho}=\sqrt{1-\rho^2}$ with $|\rho|\leq 1$, and we assume $X_0=0$ and zero interest rate without loss of generality. Note we do not have a mean reversion term λ in (1) since such a term will not materially affect the asymptotics at the leading order large deviations level that we consider in this article once we re-calibrate to the observed initial variance curve $\xi_0(t)$, but would add further headache to our already lengthy analysis in e.g. Appendix B and Lemma 2.4.

It is not known whether we have pathwise uniqueness for (1) even when $\xi_0(t)$ is constant because $\sqrt{v}$ is not Lipschitz at zero (see section 4.2.3 in [JMP20] for more on this), but we do have weak uniqueness (see Theorem 3.4 in [ALP19]) and uniqueness in law for V on $C([0,T])$, since we can explicitly compute an exponential-affine formula for the Fourier transform of V on pathspace in terms of a Volterra integral equation with a unique solution, see Appendix B (which is based on Theorem 7.1 in [ALP19]) (see also Theorem 6.1 in [ALP19]).

To clarify these points further, if we assume we have two solutions U and V to (1), then

$$\begin{aligned}\mathbb{E}((V_t-U_t)^2) &= \frac{1}{\Gamma(\alpha)^2}\mathbb{E}(\int_0^t (t-s)^{2H-1}\nu(\sqrt{V_s}-\sqrt{U_s})^2 ds) \leq \frac{1}{\Gamma(\alpha)^2}\mathbb{E}(\int_0^t (t-s)^{2H-1}\nu|V_s-U_s|ds \\ &\leq \frac{1}{\Gamma(\alpha)^2}\int_0^t (t-s)^{2H-1}\nu\mathbb{E}((V_s-U_s)^2)^{\frac{1}{2}}ds\end{aligned}$$

so $f(t):=\mathbb{E}((V_t-U_t)^2)$ satisfies

$$f(t) \leq \frac{1}{\Gamma(\alpha)^2}\int_0^t (t-s)^{2H-1}\nu\sqrt{f(s)}ds \tag{2}$$

but unfortunately there is a non-zero solution to $f(t)=\int_0^t (t-s)^{2H-1}\nu\sqrt{f(s)}ds$ in addition to the trivial zero solution (see Example 3.1.18 in [Brun17] for general $H\in(0,1)$ and for $H=\frac{1}{2}$, $f(t)=\frac{1}{4}\nu^2t^2$), so we cannot directly use a comparison principle in e.g. Appendix A.2 in [ACLP19] to assert that $f(t)\leq 0$. If however we replace the $\sqrt{v}$ coefficient in (1) with a Lipshitz function $\sigma(v)$ which agrees with $\sqrt{v}$ for $v\geq\delta>0$, this comparison theorem approach does show that we have pathwise uniqueness for V up to the hitting time of V to δ for any $\delta>0$ (see also [JMP20]). One can also adapt Lemma 4.10 in [JMP20] to show that $V_t>0$ Lebesgue a.e. even if V hits zero, and it is currently an open problem for what parameter combinations this is possible.

We let $\mathcal{F}_t=\mathcal{F}_t^{W,B}$. Then we know that $\xi_t(u):=\mathbb{E}(V_u|\mathcal{F}_t)$ is given by

$$\xi_t(u) = \xi_0(u)+\frac{\nu}{\Gamma(\alpha)}\int_0^t (u-s)^{H-\frac{1}{2}}\sqrt{V_s}dW_s$$

so

$$d\xi_t(u) = \frac{\nu}{\Gamma(\alpha)}(u-t)^{H-\frac{1}{2}}\sqrt{V_t}dW_t$$

and $\mathrm{VIX}_T^2 := -\frac{2}{\Delta}\mathbb{E}(\log\frac{S_{T+\Delta}}{S_T}|\mathcal{F}_T)=\frac{1}{\Delta}\mathbb{E}(\int_T^{T+\Delta}V_u du|\mathcal{F}_T)=\frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(u)du$.

2.1 The small-time LDP for $(\mathrm{VIX}_T^2-\mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}$

Using the stochastic Fubini theorem and Taylor's remainder theorem, we see that

$$\begin{aligned}\mathrm{VIX}_T^2 &= \frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(u)du \\ &= \frac{1}{\Delta}\int_T^{T+\Delta}(\xi_0(u)+\int_0^T\frac{\nu}{\Gamma(\alpha)}(u-s)^{H-\frac{1}{2}}\sqrt{V_s}dW_s)du \\ &= \frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(u)du + c_1\int_0^T((T+\Delta-s)^{\frac{1}{2}+H}-(T-s)^{\frac{1}{2}+H})\sqrt{V_s}dW_s \\ &= \frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(u)du + c_1\int_0^T((T+\Delta-s)^{\frac{1}{2}+H}-(T-s)^{\frac{1}{2}+H})d\tilde{X}_s\end{aligned} \tag{3}$$

where $c_1=\frac{\nu}{\Delta\Gamma(\alpha)(\frac{1}{2}+H)}$ and

$$\tilde{X}_t = \int_0^t\sqrt{V_s}dW_s$$

is the martingale component of the log stock price process X when $\rho=1$.

Then using that

$$\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(u)du = \frac{1}{\Delta}\int_0^{\Delta}\xi_0(u)du+\frac{1}{\Delta}T(\xi_0(\Delta)-\xi_0(0))+o(T) = \mathrm{VIX}_0^2+O(T)$$

we (formally) expect that

$$\frac{\mathrm{VIX}_T^2-\mathrm{VIX}_0^2}{T^{\frac{1}{2}-H}} \sim \frac{c_1\Delta^{\frac{1}{2}+H}}{T^{\frac{1}{2}-H}}\int_0^T\sqrt{V_s}dW_s = \frac{c\tilde{X}_T}{T^{\frac{1}{2}-H}} \tag{4}$$

as $T \to 0$, where

$$c \quad := \quad c_1 \Delta^{\frac{1}{2}+H} \quad = \quad \frac{\nu}{\Gamma(\alpha)\alpha} \Delta^{H-\frac{1}{2}}.$$

From the main Theorem 3.3 in [FGS21] we know that $\tilde{X}_T/T^{\frac{1}{2}-H}$ satisfies an LDP with some rate function $I^{\rho=1}(x)$ and speed T^{-2H} as $T \to 0$, so based on above we conjecture the following result, for which the full proof is given below.

Theorem 2.1 $(\mathrm{VIX}_T^2 - \mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}$ *satisfies the LDP as $T \to 0$ with speed T^{-2H} and rate function $J(x) := I^{\rho=1}(\frac{x}{c})$ where $I^{\rho=1}(x)$ is the same as $I(x)$ in Theorem 3.3 in [FGS21] for the special case when $\rho = 1$, and $J(x)$ is the Fenchel-Legendre transform of*

$$\bar{\Lambda}^{\rho=1}(cp) \quad := \quad \lim_{T\to 0} T^{2H} \log \mathbb{E}(e^{\frac{p}{T^{\alpha}}(\mathrm{VIX}_T^2 - \mathrm{VIX}_0^2)})$$

for $p \in (-\infty, \frac{p_+}{c})$ and $\bar{\Lambda}^{\rho=1}(cp) = +\infty$ otherwise, where $\bar{\Lambda}^{\rho=1}$ and p_+ are the same as $\bar{\Lambda}$ and p_+ in Theorem 3.3 in [FGS21] for the special case where the correlation ρ in [FGS21] is $+1$, and c is defined above.

The following Proposition extends and streamlines the proof of main Theorem 3.1 in [FGS21] to the case of the generalized rough Heston model in (1).

Proposition 2.2 $(X_T + \frac{1}{2}\langle X\rangle_T)/T^{\frac{1}{2}-H}$ *and $X_T/T^{\frac{1}{2}-H}$ satisfies the same LDP as $T \to 0$ as in Theorem 3.3 in [FGS21].*

Proof. Recall that $X_t + \frac{1}{2}\langle X\rangle_t$ is just the martingale component of the log stock price X_t. Then from Theorem B.1 in Appendix B, we know that

$$\mathbb{E}(e^{p(X_t+\frac{1}{2}\langle X\rangle_t)}) \quad = \quad e^{\int_0^t \xi_0(t-s)(\frac{1}{2}p^2 + p\rho\nu\psi(p,s) + \frac{1}{2}\nu^2\psi(p,s)^2)ds} \quad = \quad e^{\int_0^t \xi_0(t-s) D^{\alpha}\psi(p,s)ds}$$

for $t \in [0, T_\psi^*(p))$, where $\psi(p, .)$ satisfies the fractional Riccati VIE:

$$\psi(p,t) \quad = \quad \int_0^t c_\alpha (t-s)^{\alpha-1}(\frac{1}{2}p^2 \,+\, p\rho\nu\psi(p,s) + \frac{1}{2}\nu^2\psi(p,s)^2)ds \tag{5}$$

and $T_\psi^*(p) > 0$ is the explosion time for ψ, and (by e.g. Appendix A) this solution is unique. Then

$$\mathbb{E}(e^{\frac{p}{\varepsilon^\alpha}(\frac{1}{2}\langle X\rangle_{\varepsilon t} + X_{\varepsilon t})}) \quad = \quad e^{\int_0^{\varepsilon t} \xi_0(\varepsilon t - s)(\frac{1}{2}\frac{p^2}{\varepsilon^{2\alpha}} + \frac{p}{\varepsilon^\alpha}\rho\nu\psi(\frac{p}{\varepsilon^\alpha},s) + \frac{1}{2}\nu^2\psi(\frac{p}{\varepsilon^\alpha},s)^2)ds}$$

and

$$\begin{aligned}
\psi(\frac{p}{\varepsilon^\alpha}, \varepsilon t) \quad &= \quad \int_0^{\varepsilon t} c_\alpha(\varepsilon t - s)^{\alpha-1}(\frac{1}{2}\frac{p^2}{\varepsilon^{2\alpha}} \,+\, \frac{p}{\varepsilon^\alpha}\rho\nu\psi(\frac{p}{\varepsilon^\alpha}, s) + \frac{1}{2}\nu^2\psi(\frac{p}{\varepsilon^\alpha}, s)^2)ds \\
&= \quad \varepsilon \int_0^t c_\alpha(\varepsilon t - \varepsilon s)^{\alpha-1}(\frac{1}{2}\frac{p^2}{\varepsilon^{2\alpha}} \,+\, \frac{p}{\varepsilon^\alpha}\rho\nu\psi(\frac{p}{\varepsilon^\alpha}, \varepsilon s) + \frac{1}{2}\nu^2\psi(\frac{p}{\varepsilon^\alpha}, \varepsilon s)^2)ds
\end{aligned}$$

for $t \in [0, \frac{1}{\varepsilon}T_\psi^*(\frac{p}{\varepsilon^\alpha}))$. Then multiplying both sides by ε^α, we see that $\psi^\varepsilon(p,t) := \varepsilon^\alpha\psi(\frac{p}{\varepsilon^\alpha}, \varepsilon t)$ satisfies

$$\psi^\varepsilon(p,t) \quad = \quad \int_0^t c_\alpha(t-s)^{\alpha-1}(\frac{1}{2}p^2 \,+\, p\rho\nu\psi^\varepsilon(p,s) + \frac{1}{2}\nu^2\psi^\varepsilon(p,s)^2)ds$$

for $t \in [0, \frac{1}{\varepsilon}T_\psi^*(\frac{p}{\varepsilon^\alpha}))$, so we see that $\psi^\varepsilon(p,.)$ and $\psi(p,.)$ satisfy the same VIE, and hence are equal. Then

$$\begin{aligned}
\varepsilon^{2H} \log \mathbb{E}(e^{\frac{p}{\varepsilon^\alpha}(\frac{1}{2}\langle X\rangle_{\varepsilon t} + X_{\varepsilon t})}) \quad &= \quad \varepsilon^{2H} \log e^{\int_0^{\varepsilon t} \xi_0(\varepsilon t - s)(\frac{1}{2}\frac{p^2}{\varepsilon^{2\alpha}} + \frac{p}{\varepsilon^\alpha}\rho\nu\psi(\frac{p}{\varepsilon^\alpha},s) + \frac{1}{2}\nu^2\psi(\frac{p}{\varepsilon^\alpha},s)^2)ds} \\
&= \quad \varepsilon^{2H} \log e^{\varepsilon \int_0^t \xi_0(\varepsilon t - \varepsilon s)(\frac{1}{2}\frac{p^2}{\varepsilon^{2\alpha}} + \frac{p}{\varepsilon^\alpha}\rho\nu\psi(\frac{p}{\varepsilon^\alpha},\varepsilon s) + \frac{1}{2}\nu^2\psi(\frac{p}{\varepsilon^\alpha},\varepsilon s)^2)ds} \\
&= \quad \int_0^t \xi_0(\varepsilon t - \varepsilon s)(\frac{1}{2}p^2 \,+\, p\rho\nu\psi(p,s) + \frac{1}{2}\nu^2\psi(p,s)^2)ds \\
&\to \quad V_0 \int_0^t (\frac{1}{2}p^2 \,+\, p\rho\nu\psi(p,s) + \frac{1}{2}\nu^2\psi(p,s)^2)ds \quad = \quad \bar{\Lambda}(p,t)
\end{aligned} \tag{6}$$

as $\varepsilon \to 0$ if $t < T_\psi^*(p)$ using the bounded convergence theorem, since $\xi_0(.)$ is continuous at zero and ψ is bounded on $[0,t]$ for if $t < T_\psi^*(p)$. From Lemma 2.3.9 in [DZ98], we know that $\bar{\Lambda}(p,t)$ is convex in p, and from (5) we also know that

$$\frac{d}{dt}\Lambda(p,t) \quad = \quad \frac{1}{2}p^2 \,+\, p\rho\nu\psi(p,t) \,+\, \frac{1}{2}\nu^2\psi(p,t)^2$$

so $\Lambda(p,t)$ is differentiable in t.

Using the scaling relation in Corollary 3.4 in [FGS21] we also know that $\Lambda(p,1) = p^{\frac{2H}{\alpha}}\Lambda(\mathrm{sgn}(p),|p|^{\frac{1}{\alpha}})$, so setting $\Lambda(p) := \Lambda(p,1)$ we know that $\Lambda(p)$ is differentiable in p. Moreover, the quadratic $Q(w) := \frac{1}{2}p^2 + \rho p\nu w^2 + \frac{1}{2}w^2$ has no real roots so we are in Case A or B in [GGP19] where the VIE for $\psi(p,.)$ has no fixed point, so T^*_ψ is finite and explodes at rate $const./(T^*_\psi(p)-t)^\alpha$ (see Lemma 3 in [GGP19]).

From the integral in (6) and the aforementioned known explosion rate and the scaling relation, we see that $\bar{\Lambda}(p,t)$ also tends to $+\infty$ as $p \nearrow p_+ = T^*_\psi(+1)^\alpha$ or as $p \searrow p_- = -T^*_\psi(-1)^\alpha$, and (by convexity and differentiability) Λ is also essentially smooth. Moreover, from the monotonicity of the L^p-norm, we know that $\bar{\Lambda}(p,t) = \infty$ for $p \notin (p_-,p_+)$ as well. Hence by the Gärtner-Ellis theorem from large deviations theory (see Theorem 2.3.6 in [DZ98]), $(X_\varepsilon + \frac{1}{2}\langle X\rangle_\varepsilon)/\varepsilon^{\frac{1}{2}-H}$ satisfies the LDP as $\varepsilon \to 0$ with speed ε^{-2H} and rate function $I(x)$. Finally the LDP for $X_\varepsilon/\varepsilon^{\frac{1}{2}-H}$ is obtained using exponential equivalence as in [FGS21]. ■

Proof. (of Theorem 2.1). Setting $T=\varepsilon$ to make the notation consistent with [FGS21] and integrating (3) by parts, we see that

$$\begin{aligned}
\mathrm{VIX}^2_\varepsilon - \frac{1}{\Delta}\int_\varepsilon^{\varepsilon+\Delta}\xi_0(u)du \;=\; \mathrm{VIX}^2_\varepsilon - \mathrm{VIX}^2_0 - \zeta \;&=\; c_1\int_0^\varepsilon((\varepsilon+\Delta-s)^{\frac{1}{2}+H} - (\varepsilon-s)^{\frac{1}{2}+H})d\tilde{X}_s &(7)\\
&=\; c_1\Delta^{\frac{1}{2}+H}\tilde{X}_\varepsilon - c_1\alpha\int_0^\varepsilon((\varepsilon-s)^{H-\frac{1}{2}} - (\varepsilon+\Delta-s)^{H-\frac{1}{2}})\tilde{X}_s ds\\
&=\; c_1\Delta^{\frac{1}{2}+H}\tilde{X}_\varepsilon + \int_0^\varepsilon f(\varepsilon-s)\tilde{X}_s ds\\
&=\; c\tilde{X}_\varepsilon + (f*\tilde{X})_\varepsilon &(8)
\end{aligned}$$

where $\zeta := \frac{1}{\Delta}\int_\varepsilon^{\varepsilon+\Delta}\xi_0(u)du - \mathrm{VIX}^2_0 = O(\varepsilon)$ and $f(s) := c_1\alpha((s+\Delta)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}})$ and we note that $f \in L^2$. As discussed above, from the main Theorem 3.1 in [FGS21] we know that the leading order term $c\tilde{X}_T/T^{\frac{1}{2}-H}$ satisfies the stated LDP as $T \to 0$, so the issue is just to argue away the remainder term, using exponential equivalance as in [FGS21].

From Theorem B.1 (which is adapted from Eqs 2.8-2.10 in [AE19] and Lemma 7.3 in [ALP19]), we know that

$$\mathbb{E}(e^{p\int_0^\varepsilon f(\varepsilon-s)\tilde{X}_s ds}) \;=\; e^{\int_0^\varepsilon \xi_0(\varepsilon-s)D^\alpha\psi_2(p,s)ds} \;=\; e^{\int_0^\varepsilon \xi_0(\varepsilon-s)g(p,s)ds}$$

where

$$\begin{aligned}
\psi_1(p,t) \;&=\; p\int_0^t f(s)ds\\
\psi_2(p,t) \;&=\; \int_0^t c_\alpha(t-s)^{\alpha-1}(\frac{1}{2}\psi_1(p,s)^2 + \psi_1(p,s)\nu\psi_2(p,s) + \frac{1}{2}\nu^2\psi_2(p,s)^2)ds &(9)
\end{aligned}$$

for $\varepsilon \le T_{\psi_2}(p)$ where $T_{\psi_2}(p)$ is the explosion time for ψ_2 (note that $g(p,t) := D^\alpha\psi_2(p,t) = \frac{1}{2}\psi_1(p,t)^2 + \psi_1(p,t)\nu\psi_2(p,t) + \frac{1}{2}\nu^2\psi_2(p,t)^2)$, and recall that $f(s) := c_1\alpha((s+\Delta)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}})$. We first note that

$$\psi_1(\varepsilon^{-\alpha}p,\varepsilon t) \;=\; \frac{p}{\varepsilon^{H+\frac{1}{2}}}c_1\alpha\int_0^{\varepsilon t}((s+\Delta)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}})ds \;=\; pc\varepsilon^{-\alpha}h_\varepsilon(t) \qquad (10)$$

where h_ε denotes the bounded, continuous function

$$h_\varepsilon(t) \;:=\; \Delta^{-\frac{1}{2}-H}\alpha\int_0^{\varepsilon t}((s+\Delta)^{H-\frac{1}{2}} - s^{H-\frac{1}{2}})ds \;\le\; 0$$

defined for $t\in[0,1]$, which tends to zero pointwise as $\varepsilon \to 0$ (this will be needed below). Then

$$\begin{aligned}
\psi_2(p\varepsilon^{-\alpha},\varepsilon t) \;&=\; \int_0^{\varepsilon t} c_\alpha(\varepsilon t-s)^{\alpha-1}(\frac{1}{2}\psi_1(p\varepsilon^{-\alpha},s)^2 + \psi_1(p\varepsilon^{-\alpha},s)\nu\psi_2(p\varepsilon^{-\alpha},s) + \frac{1}{2}\nu^2\psi_2(p\varepsilon^{-\alpha},s)^2)ds\\
&=\; \varepsilon\int_0^t c_\alpha(\varepsilon t-\varepsilon s)^{\alpha-1}(\frac{1}{2}\psi_1(p\varepsilon^{-\alpha},\varepsilon s)^2 + \psi_1(p\varepsilon^{-\alpha},\varepsilon s)\rho\nu\psi_2(p\varepsilon^{-\alpha},\varepsilon s) + \frac{1}{2}\nu^2\psi_2(p\varepsilon^{-\alpha},\varepsilon s)^2)ds\\
&=\; \varepsilon^\alpha\int_0^t c_\alpha(t-s)^{\alpha-1}(\frac{1}{2}(pc\varepsilon^{-\alpha}h_\varepsilon(s))^2 + pc\varepsilon^{-\alpha}h_\varepsilon(s)\nu\psi_2(p\varepsilon^{-\alpha},\varepsilon s) + \frac{1}{2}\nu^2\psi_2(p\varepsilon^{-\alpha},\varepsilon s)^2)ds
\end{aligned}$$

for $t\in[0,\frac{1}{\varepsilon}T_{\psi_2}(\varepsilon^{-\alpha}p))$. Multiplying by ε^α and cancelling powers of ε, we see that

$$\psi_2^\varepsilon(p,t) \;:=\; \varepsilon^\alpha\psi_2(p\varepsilon^{-\alpha},\varepsilon t) \;=\; \int_0^t c_\alpha(t-s)^{\alpha-1}(\frac{1}{2}(pch_\varepsilon(s))^2 + pch_\varepsilon(s)\rho\nu\psi_2^\varepsilon(p,s) + \frac{1}{2}\nu^2\psi_2^\varepsilon(p,s)^2)ds$$

i.e. $\psi_2^\varepsilon(p,t)$ satisfies

$$D^\alpha \psi_2^\varepsilon(p,t) \;=\; \frac{1}{2}(pch_\varepsilon(t)+\nu\psi_2^\varepsilon(p,t))^2\,. \tag{11}$$

Then we see that

$$\begin{aligned}
\varepsilon^{2H}\log\mathbb{E}(e^{p\varepsilon^{-\alpha}(\mathrm{VIX}_\varepsilon^2-\zeta-\mathrm{VIX}_0^2-c\tilde{X}_\varepsilon)}) &= \varepsilon^{2H}\log\mathbb{E}(e^{p\varepsilon^{-\alpha}\int_0^\varepsilon f(\varepsilon-s)\tilde{X}_s ds})\\
&= \varepsilon^{2H}\log e^{\int_0^\varepsilon \xi_0(\varepsilon-s)g(p\varepsilon^{-\alpha},s)ds}\\
&= \varepsilon^{2H}\log e^{\varepsilon\int_0^1 \xi_0(\varepsilon-\varepsilon u)g(p\varepsilon^{-\alpha},\varepsilon u)du}\\
&= \varepsilon^{2\alpha}\int_0^1 \xi_0(\varepsilon-\varepsilon u)g(p\varepsilon^{-\alpha},\varepsilon u)du\\
&= \int_0^1 \xi_0(\varepsilon-\varepsilon u)g_\varepsilon(p,u)du\\
&= \int_0^1 \xi_0(\varepsilon-\varepsilon u)\frac{1}{2}(pch_\varepsilon(s)+\nu\psi_2^\varepsilon(p,s))^2 ds\,.
\end{aligned}$$

where $g_\varepsilon(p,u):=g(p\varepsilon^{-\alpha},\varepsilon u)$.

Recall that $h_\varepsilon(t)\to 0$ as $\varepsilon\to 0$, so we expect $\psi^\varepsilon(t)$ to tend to zero as well. To use Theorem 13.1.1 in [GLS90] to prove this, we first need to verify uniqueness for the solution ψ_ε, which we can do using the general argument given in Appendix A.

Since (11) with h_ε replaced by zero has a unique solution equal to the zero function, from Theorem 13.1.1 i) in [GLS90] we know there is a subsequence ε_n such that $\psi_{\varepsilon_n}(p,t)$ converges uniformly to zero on $[0,1]$. Now suppose $\psi^\varepsilon(p,.)$ does not converge uniformly to zero. Then we can find a subsequence ε_k such that $\psi_{\varepsilon_k}(p,.)$ stays uniformly far from zero for all $k\in\mathbb{N}$. This subsequence has no subsequence that converges to zero, which contradicts Theorem 13.1.1 i) in [GLS90].

Then using (11) and the bounded convergence theorem and the continuity of $\xi_0(t)$ at $t=0$ we see that

$$\lim_{\varepsilon\to 0}\varepsilon^{2H}\log\mathbb{E}(e^{p\varepsilon^{-\alpha}(\mathrm{VIX}_\varepsilon^2-\zeta-\mathrm{VIX}_0^2-c\tilde{X}_\varepsilon)}) \;=\; \lim_{\varepsilon\to 0}\int_0^1 \xi_0(\varepsilon-\varepsilon u)D^\alpha g_\varepsilon(p,u)du \;=\; 0 \tag{12}$$

for all $p\in\mathbb{R}$, and since $\zeta=O(\varepsilon)$ the limit is unchanged if we remove ζ here. Finally, setting $R_T:=\mathrm{VIX}_T^2-\mathrm{VIX}_0^2-c\tilde{X}_T$, and using (12) we see that for $x>0$ and $p>0$

$$\lim_{T\to 0}T^{2H}\log\,\mathbb{P}(\frac{R_T}{T^{\frac{1}{2}-H}}>x) \;\leq\; \lim_{T\to 0}T^{2H}\log\,\mathbb{E}(e^{\frac{p}{T^{2H}}(\frac{R_T}{T^{\frac{1}{2}-H}}-x)}) \;=\; 0-xp$$

and taking the inf over $p\geq 0$ we see that the left hand side is $-\infty$. Similarly for $x<0$ and $p<0$

$$\lim_{T\to 0}T^{2H}\log\,\mathbb{P}(\frac{R_T}{T^{\frac{1}{2}-H}}<x) \;\leq\; \lim_{T\to 0}T^{2H}\log\,\mathbb{E}(e^{\frac{p}{T^{2H}}(\frac{R_T}{T^{\frac{1}{2}-H}}-x)}) \;=\; -xp$$

and again we can take the inf over $p\leq 0$. Combining these observations, we see that

$$\lim_{T\to 0}T^{2H}\log\,\mathbb{P}(|\frac{R_T}{T^{\frac{1}{2}-H}}|>x) \;=\; -\infty$$

which shows that $(\mathrm{VIX}_T^2-\mathrm{VIX}_0^2-\zeta)/T^{\frac{1}{2}-H}$ and $c\tilde{X}_T/T^{\frac{1}{2}-H}$ are exponentially equivalent as $T\to 0$ (where ζ is defined in (7)) (see Definition 4.2.10 in [DZ98]), so the LDP follows from Theorem 4.2.13 in [DZ98] (as used in [FGS21]). Finally we can remove the ζ term here since ζ is deterministic and $o(T^{\frac{1}{2}-H})$ so $(\mathrm{VIX}_T^2-\mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}$ and $c\tilde{X}_T/T^{\frac{1}{2}-H}$ are also exponentially equivalent. ■

Remark 2.1 The lower bound for p in Theorem 2.1 is $-\infty$ (as opposed to some finite negative constant p_-) because for $\rho=1$ and $p<0$, $\bar{\Lambda}^{\rho=1}(.)$ falls under case C for the ABCD classification used in [GGP19] (for our case we have to use driftless versions of the quantities defined in Eq 7 and 8 in [GGP19]; specifically $c_1(u)=\frac{1}{2}u^2$, $e_0(u)=\frac{1}{2}\rho\nu u$ and $e_1(u)=e_0(u)^2-\frac{1}{4}\nu^2u^2$, because we are working with $\tilde{X}$ not the true log stock price process X). But since the ρ value associated with the $\tilde{X}$ process is 1, we are in the special double root case for Eq 10 in [GGP19] where $e_1(u)=0$ (borderline between C and B), but since we still fall in Case C, there is no explosion for the VIE in Eq 24 in [FGS21] for any negative p-value.

Remark 2.2 For the driftless case where $\xi_0(t)\equiv V_0$, using a simple ansatz and local martingale arguments, Proposition 4.6 in [GK19] derives the following exponential-affine formula for the mgf of $\int_T^{T+\Delta}\xi_T(u)du$:

$$\mathbb{E}(e^{h\int_T^{T+\Delta}(\xi_T(u)-V_0)du}) \;=\; e^{V_0\int_0^T g(T+\Delta-u)du} \;=\; e^{V_0\int_\Delta^{\Delta+T}g(s)ds}$$

for h in a certain interval, where g satisfies the non-standard VIE $g(t)=\frac{1}{2}(\int_0^t \frac{\nu}{\Gamma(\alpha)}(t-v)^{H-\frac{1}{2}}g(v)dv)^2$ for $t\geq\Delta$ and $g(t)=h$ for $t\in[0,\Delta]$ (note g is discontinuous at $t=\Delta$ or else we have a contradiction). This is clearly very relevant for pricing VIX options at non-zero maturities using Fourier inversion methods (see Subsection 2.4 for more details), but we will not need to use this VIE in this article.

2.2 VIX call option asymptotics

We now translate the LDP in Theorem 2.1 into small-time asymptotics for VIX call options for the same large deviations regime used in [FGS21]:

Corollary 2.3 *For $x>0$ we have the following asymptotic behaviour for close-to-the money VIX call option prices:*

$$\lim_{T\to 0} T^{2H}\log\mathbb{E}((\mathrm{VIX}_T-\mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}})_+) \;=\; -J(2\mathrm{VIX}_0^2 x)$$

where J is the rate function defined in the main Theorem 2.1.

Proof. See Appendix D. ∎

Remark 2.3 For $x<0$ (using very similar arguments), we obtain the following small-time behaviour for close-to-the-money VIX put options:

$$\lim_{T\to 0} T^{2H}\log\mathbb{E}((\mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}}-\mathrm{VIX}_T)_+) \;=\; -J(2\mathrm{VIX}_0^2 x)\,.$$

2.3 VIX future and implied volatility asymptotics

Lemma 2.4 $\frac{\mathrm{VIX}_T^2-\mathrm{VIX}_0^2}{\sqrt{T}}$ *tends weakly to $c\sqrt{V_0}Z$, where Z is a standard Normal.*

Proof. From Theorem B.1 (which is adapted from Eqs 2.8-2.10 in [AE19] and Lemma 7.3 in [ALP19]), we know that

$$\mathbb{E}(e^{pc\tilde{X}_\varepsilon+p\int_0^\varepsilon f(\varepsilon-s)\tilde{X}_s ds}) \;=\; e^{\int_0^\varepsilon \xi_0(\varepsilon-s)D^\alpha\psi_2(p,s)ds} \;=\; e^{\int_0^\varepsilon \xi_0(\varepsilon-s)g(p,s)ds}$$

where

$$\begin{aligned}\psi_1(p,t) &= p(c+\int_0^t f(s)ds)\\ \psi_2(p,t) &= \int_0^t c_\alpha(t-s)^{\alpha-1}(\frac{1}{2}\psi_1(p,s)^2+\psi_1(p,s)\nu\psi_2(p,s)+\frac{1}{2}\nu^2\psi_2(p,s)^2)ds\end{aligned} \tag{13}$$

for $\varepsilon\leq T^*_{\psi_2}(p)$ where $T^*_{\psi_2}(p)$ is the explosion time for ψ_2 (where $g(p,t):=D^\alpha\psi_2(p,t)=\frac{1}{2}\psi_1(p,t)^2+\psi_1(p,t)\rho\nu\psi_2(p,t)+\frac{1}{2}\nu^2\psi_2(p,t)^2)$, and recall that $f(s):=c_1\alpha((s+\Delta)^{H-\frac{1}{2}}-s^{H-\frac{1}{2}})$. We first note that

$$\psi_1(\varepsilon^{-\frac{1}{2}}p,\varepsilon t) \;=\; p\varepsilon^{-\frac{1}{2}}c+\frac{p}{\varepsilon^{\frac{1}{2}}}c_1\alpha\int_0^{\varepsilon t}((s+\Delta)^{H-\frac{1}{2}}-s^{H-\frac{1}{2}})ds \;=\; pc\varepsilon^{-\frac{1}{2}}(1+h_\varepsilon(t)) \tag{14}$$

where h_ε is defined as above. Then

$$\begin{aligned}\psi_2(p\varepsilon^{-\frac{1}{2}},\varepsilon t) &= \int_0^{\varepsilon t} c_\alpha(\varepsilon t-s)^{\alpha-1}(\frac{1}{2}\psi_1(p\varepsilon^{-\frac{1}{2}},s)^2+\psi_1(p\varepsilon^{-\frac{1}{2}},s)\nu\psi_2(p\varepsilon^{-\frac{1}{2}},s)+\frac{1}{2}\nu^2\psi_2(p\varepsilon^{-\frac{1}{2}},s)^2)ds\\ &= \varepsilon\int_0^t c_\alpha(\varepsilon t-\varepsilon s)^{\alpha-1}(\frac{1}{2}\psi_1(p\varepsilon^{-\frac{1}{2}},\varepsilon s)^2+\psi_1(p\varepsilon^{-\frac{1}{2}},\varepsilon s)\nu\psi_2(p\varepsilon^{-\frac{1}{2}},\varepsilon s)+\frac{1}{2}\nu^2\psi_2(p\varepsilon^{-\frac{1}{2}},\varepsilon s)^2)ds\\ &= \varepsilon^\alpha\int_0^t c_\alpha(t-s)^{\alpha-1}(\frac{1}{2}(pc\varepsilon^{-\frac{1}{2}}(1+h_\varepsilon(s)))^2+pc\varepsilon^{-\frac{1}{2}}(1+h_\varepsilon(s))\nu\psi_2(p\varepsilon^{-\frac{1}{2}},\varepsilon s)+\frac{1}{2}\nu^2\psi_2(p\varepsilon^{-\frac{1}{2}},\varepsilon s)^2)ds\end{aligned}$$

for $t\in[0,\frac{1}{\varepsilon}T_{\psi_2}(\varepsilon^{-\frac{1}{2}}p))$. Multiplying by $\sqrt{\varepsilon}$ and cancelling powers of ε, we see that $\psi_2^\varepsilon(p,t):=\sqrt{\varepsilon}\psi_2(p\varepsilon^{-\frac{1}{2}},t)$ satisfies

$$\psi_2^\varepsilon(p,t) \;:=\; \varepsilon^H\int_0^t c_\alpha(t-s)^{\alpha-1}(\frac{1}{2}(pc(1+h_\varepsilon(s))^2+pc(1+h_\varepsilon(s))\nu\psi_2^\varepsilon(p,s)+\frac{1}{2}\nu^2\psi_2^\varepsilon(p,s)^2)ds$$

i.e. $\psi_2^\varepsilon(p,t)$ satisfies $D^\alpha\psi_2^\varepsilon(p,t)=\frac{1}{2}\varepsilon^H(pc(1+h_\varepsilon(t))+\nu\psi_2^\varepsilon(p,t))^2$. Then we see that

$$\begin{aligned}\mathbb{E}(e^{p\varepsilon^{-\frac{1}{2}}(\mathrm{VIX}_\varepsilon^2-\zeta-\mathrm{VIX}_0^2)}) &= e^{\int_0^\varepsilon \xi_0(\varepsilon-s)\frac{1}{2}(\psi_1(\frac{p}{\sqrt{\varepsilon}},s)+\nu\psi_2(\frac{p}{\sqrt{\varepsilon}},s))^2ds}\\ &= e^{\varepsilon\int_0^1 \xi_0(\varepsilon-\varepsilon s)\frac{1}{2}(\psi_1(\frac{p}{\sqrt{\varepsilon}},\varepsilon s)+\nu\psi_2(\frac{p}{\sqrt{\varepsilon}},\varepsilon s))^2ds}\\ &= e^{\varepsilon\int_0^1 \xi_0(\varepsilon-\varepsilon s)\frac{1}{2}(pc\varepsilon^{-\frac{1}{2}}(1+h_\varepsilon(t))+\nu\psi_2(\frac{p}{\sqrt{\varepsilon}},\varepsilon s))^2ds}\\ &= e^{\int_0^1 \xi_0(\varepsilon-\varepsilon s)\frac{1}{2}(pc(1+h_\varepsilon(t))+\nu\psi_2^\varepsilon(p,s))^2ds}. \qquad (15)\end{aligned}$$

ε^H and $h_\varepsilon(t)\to 0$ as $\varepsilon\to 0$, so we expect $\psi_2^\varepsilon(t)$ to tend to zero as well. To use Theorem 13.1.1 in [GLS90] to prove this, we first need to verify uniqueness for the solution ψ^ε, which we can do using the general argument given in Appendix A.

From Theorem 13.1.1 i) in [GLS90] (as above) we know that $\psi_2^\varepsilon(p,.)$ converges uniformly to zero on any compact interval, and $\psi_2^\varepsilon(p,t)$ is continuous in ε and t on $\{(\varepsilon,t):\varepsilon\in[0,1),0\le t<T_\varepsilon^*(p)\}$. Then using (15) and the bounded convergence theorem and the continuity of $\xi_0(t)$ at $t=0$ we see that

$$\lim_{\varepsilon\to 0}\mathbb{E}(e^{p\varepsilon^{-\frac{1}{2}}(\mathrm{VIX}_\varepsilon^2-\zeta-\mathrm{VIX}_0^2)}) = e^{\frac{1}{2}p^2V_0c^2}$$

for all $p\in\mathbb{R}$, and since $\zeta=O(\varepsilon)$ the limit is unchanged if we remove ζ here. Finally, since Theorem 13.1.1 in [GLS90] is multi-dimensional, we can apply it to $(\mathrm{Re}(\psi),\mathrm{Im}(\psi))$ with p replaced by ik with $k\in\mathbb{R}$ as we do in section 5 in [FGS21]. The result then follows from Lévy's convergence theorem. ■

We also have the following asymptotic estimate for the small-time behaviour of VIX futures which will be needed for the implied volatility asymptotics below.

Lemma 2.5 $\mathbb{E}(\mathrm{VIX}_T-\mathrm{VIX}_0)=O(\sqrt{T})$ *as* $T\to 0$.

Proof. Recall that $\mathrm{VIX}_\varepsilon^2-\mathrm{VIX}_0^2-\zeta=c\tilde{X}_\varepsilon+(f*\tilde{X})_\varepsilon$ from (8) where $\zeta=O(\varepsilon)$ and $f(s):=c_1\alpha((s+\Delta)^{H-\frac{1}{2}}-s^{H-\frac{1}{2}})$ and setting $d\hat{X}_t:=\int_0^t\sqrt{V_0}(\rho dW_s+\bar{\rho}dB_s)$ and $\varepsilon=t$, we see that

$$\begin{aligned}\mathbb{E}((\mathrm{VIX}_t^2-\mathrm{VIX}_0^2-\zeta)^2)^{\frac{1}{2}} &= \mathbb{E}((c\tilde{X}_t+(f*\tilde{X})_t)^2)^{\frac{1}{2}}\\ &\le \mathbb{E}([c(\tilde{X}_t-\hat{X}_t)+(f*(\tilde{X}-\hat{X}))_t]^2)^{\frac{1}{2}}+\mathbb{E}((c\hat{X}_t+(f*\hat{X})_t)^2)^{\frac{1}{2}}\\ &\le c\mathbb{E}((\tilde{X}_t-\hat{X}_t)^2)^{\frac{1}{2}}+\mathbb{E}((f*(\tilde{X}-\hat{X}))_t^2)^{\frac{1}{2}}+c\sqrt{V_0}\sqrt{t}\\ &+ \mathbb{E}((\int_0^t f(t-s)\int_0^s\sqrt{V_0}dW_u ds)^2)^{\frac{1}{2}}.\end{aligned}$$

Using stochastic Fubini we can re-write the final term as $\mathbb{E}((\sqrt{V_0}\int_0^t\int_u^t f(t-s)dsdW_u)^2)^{\frac{1}{2}}=O(t)$ [2]. We also note that

$$\begin{aligned}\mathbb{E}((\tilde{X}_t-\hat{X}_t)^2) = \int_0^t\mathbb{E}((\sqrt{V_s}-\sqrt{V_0})^2)ds &= \int_0^t(\xi_0(s)+2\sqrt{V_0}\mathbb{E}(\sqrt{V_s})+V_0)ds\\ &\le \int_0^t(\xi_0(s)+2\sqrt{V_0}\mathbb{E}(V_s)^{\frac{1}{2}}+V_0)ds\\ &\le \int_0^t(\xi_0(s)+2\sqrt{V_0}\xi_0(s)^{\frac{1}{2}}+V_0)ds \sim 4V_0t \qquad (16)\end{aligned}$$

as $t\to 0$, and for the convolution term (from Jensen) we see that

$$\mathbb{E}((f*(\tilde{X}-\hat{X}))_t^2)^{\frac{1}{2}} = \mathbb{E}((t\cdot\frac{1}{t}\int_0^t f(t-s)(\tilde{X}_s-\hat{X}_s)ds)^2)^{\frac{1}{2}} \le t\int_0^t f(t-s)^2\mathbb{E}((\tilde{X}_s-\hat{X}_s)^2)ds = O(t).$$

using (16) and the fact that $f\in L^2$.

Putting all this together, we see that $\frac{1}{\sqrt{t}}\mathbb{E}((\mathrm{VIX}_t^2-\mathrm{VIX}_0^2-\zeta)^2)^{\frac{1}{2}}\le\bar{c}$ for some constant $\bar{c}>0$ and t sufficiently small, and since $\zeta=O(t)$ we can remove the ζ term and the result still holds. Thus $\Upsilon_T:=\frac{\mathrm{VIX}_T^2-\mathrm{VIX}_0^2}{\sqrt{T}}$ is U.I., and (from Lemma 2.4) we know that $\Upsilon_T\overset{w}{\to}c\sqrt{V_0}Z$ as $\varepsilon\to 0$, where $Z\sim N(0,1)$.

Then from (7) and the Ito isometry we know that

$$\begin{aligned}\mathbb{E}((\mathrm{VIX}_T^2-\mathrm{VIX}_0^2)^2)^{\frac{1}{2}} &\le \mathbb{E}((\mathrm{VIX}_T^2-\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(u)du)^2)^{\frac{1}{2}}+|\frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(u)du-\mathrm{VIX}_0^2|\\ &= c_1(\int_0^T((T+\Delta-s)^{\frac{1}{2}+H}-(T-s)^{\frac{1}{2}+H})^2\xi_0(s)ds)^{\frac{1}{2}}+|\zeta| \to 0 \qquad (17)\end{aligned}$$

[2] can easily check this in Mathematica

as $T \to 0$, so $\mathrm{VIX}_T^2 \to \mathrm{VIX}_0^2$ in L^2 and hence also in probability.

Now define $Y_T := \frac{1}{\mathrm{VIX}_T+\mathrm{VIX}_0}$. Then Y_T is a continuous function of VIX_T^2 so (by the continuous mapping theorem) $Y_T \to Y_0$ (a constant) in probability, and clearly $Y_T \leq \frac{1}{\mathrm{VIX}_0}$. Note that $\frac{\mathrm{VIX}_T-\mathrm{VIX}_0}{\sqrt{T}} = \Upsilon_T Y_T$, and from above we know that $\Upsilon_T \overset{w}{\to} c\sqrt{V_0}Z$. From the general standard result that if $X_n \overset{w}{\to} X$ and $Y_n \to c$ (a constant) in probability, then $(X_n, Y_n) \overset{w}{\to} (X, c)$, we see that (Υ_T, Y_T) tends weakly to $(c\sqrt{V_0}Z, Y_0)$, and from the continuous mapping theorem $\Upsilon_T Y_T$ tends weakly to $Y_0 Z$. Moreover, Y_T is uniformly bounded so $\Upsilon_T Y_T$ is also U.I. Then by Theorem 3.5 in Billingsley[Bil99], $\mathbb{E}(\Upsilon_T Y_T) \to Y_0\mathbb{E}(Z) = 0$. ■

Corollary 2.6 *If $\hat{\sigma}_{\mathrm{VIX}}(K,T)$ denotes the implied volatility of a VIX call or put option with strike K, we see that*

$$\hat{\sigma}_{\mathrm{VIX}}(x) \;:=\; \lim_{T\to 0}\hat{\sigma}_{\mathrm{VIX}}(\mathrm{VIX}_0 e^{xT^{H-\frac{1}{2}}}, T) \;=\; \frac{|x|}{\sqrt{2J(2\mathrm{VIX}_0^2 x)}} \tag{18}$$

for $x \in \mathbb{R}$, where J is the rate function introduced in the main Theorem 2.1.

Proof. Let $C^{\mathrm{BS}}(S,K,\sigma,T)$ denote the usual Black-Scholes call option formula with zero interest rate and dividend. Then can easily verify that for any $b \in \mathbb{R}$

$$\lim_{T\to 0} T^{2H}\log C^{\mathrm{BS}}(\mathrm{VIX}_0 + b\sqrt{T}, \mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}}, \sigma, T) \;=\; -\frac{x^2}{2\sigma^2}$$

so from Lemma 2.5 (and using that C^{BS} is monotonic in its first argument) we see that

$$\lim_{T\to 0} T^{2H}\log C^{\mathrm{BS}}(\mathbb{E}(\mathrm{VIX}_T), \mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}}, \sigma, T) \;=\; -\frac{x^2}{2\sigma^2}\,.$$

For any $\delta \in (0, J(2\mathrm{VIX}_0^2 x))$, we can then choose σ so that $-J(2\mathrm{VIX}_0^2 x) = -\frac{x^2}{2\sigma^2} - \delta$. Then from Corollary 2.3

$$\begin{aligned}
-J(2\mathrm{VIX}_0^2 x) &= \limsup_{T\to 0} T^{2H}\log \mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}})^+) \\
&= \limsup_{T\to 0} T^{2H}\log C^{\mathrm{BS}}(\mathbb{E}(\mathrm{VIX}_T), \mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}}, \hat{\sigma}_{\mathrm{VIX}}(x,T), T) \quad \text{(by definition of } \hat{\sigma}_{\mathrm{VIX}}(x,T)) \\
&< \lim_{T\to 0} T^{2H}\log C^{\mathrm{BS}}(\mathbb{E}(\mathrm{VIX}_T), \mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}}, \sigma, T) \;=\; -\frac{x^2}{2\sigma^2}\,.
\end{aligned}$$

Since $C^{\mathrm{BS}}(.)$ is monotonically increasing in the σ argument, we see that $\limsup_{T\to 0}\hat{\sigma}_{\mathrm{VIX}}(x,T) \leq \sigma$. Finally we let $\delta \to 0$, and we proceed similarly for the lower bound. ■

2.3.1 Small log-moneyness expansions

Using section 3.4 in [FGS21], we obtain the following small-moneyness expansion

$$\begin{aligned}
\bar{\Lambda}^{\rho=1}(p) &= \frac{1}{2}V_0 p^2 + \frac{V_0\nu}{2\Gamma(2+\alpha)}p^3 + O(p^3) \\
(\bar{\Lambda}^{\rho=1})^*(x) &= \frac{1}{2}\frac{x^2}{V_0} - \frac{\nu x^3}{2V_0^2\Gamma(2+\alpha)} + O(x^4)
\end{aligned}$$

and combining this with (18) we find that

$$\begin{aligned}
\hat{\sigma}_{\mathrm{VIX}}(x) &= \frac{\bar{c}\nu\sqrt{V_0}}{2\mathrm{VIX}_0^2} + \frac{\nu x}{2\sqrt{V_0}\Gamma(2+\alpha)} \\
&+ \frac{1}{V_0^{\frac{3}{2}}}\mathrm{VIX}_0^2\nu\alpha\Delta^{1-\alpha}\Gamma(\alpha)\frac{\Gamma(2+\alpha)^2\Gamma(1+2\alpha)+\Gamma(1+\alpha)^2(4\Gamma(2+\alpha)^2-6\Gamma(2+2\alpha))}{4\Gamma(1+\alpha)^2\Gamma(2+\alpha)^2\Gamma(2+2\alpha)}x^2 + O(x^3)
\end{aligned} \tag{19}$$

where $\bar{c} := \frac{1}{\Gamma(\alpha)\alpha}\Delta^{\alpha-1}$ and we see that the linear skew term is positive, and note there is no VIX smile if $\nu = 0$, since in this case V_t is constant. Moreover, since the fraction in front of the x^2 term only depends on α, we can readily verify from a graph that the $O(x^2)$ convexity term is strictly negative (see Figure 1 below), which is consistent with what is observed in practice, see e.g. [JMP21] and plots in [GJR20],[Guy20],[DeM18],[HJT20] et al. for more on this point. Since V_0 is already fixed from $\xi_0()$ i.e. $V_0 = \xi_0(0)$ we see that we cannot independently fit the overall level and the skew of the VIX smile in the small-T limit. This issue is addressed in the companion article [FS21] by the addition of an independent CGMY-jump component to the model which allows the SPX and VIX short-maturity smiles to decouple in some sense.

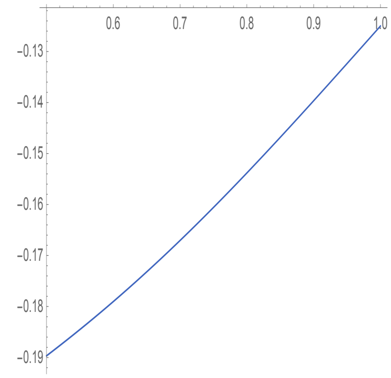


Figure 1: Here we have plotted the factor $\frac{\Gamma(2+\alpha)^2\Gamma(1+2\alpha)+\Gamma(1+\alpha)^2(4\Gamma(2+\alpha)^2-6\Gamma(2+2\alpha))}{4\Gamma(1+\alpha)^2\Gamma(2+\alpha)^2\Gamma(2+2\alpha)}$ which appears in the convexity term in (19) as a function of α, and we see that this factor is strictly negative for all admissible α values.

2.3.2 The Edgeworth regime

Proceeding as in [FSV21], we have also formally verified the following asymptotic behaviour for VIX options in the Edgeworth regime under driftless rough Heston model :

$$\hat{\sigma}_{\mathrm{VIX}}(\sqrt{V_0}e^{x\sqrt{T}},T) \quad = \quad \frac{\nu}{\sqrt{V_0}}(\frac{\Delta^{\alpha-1}}{2\alpha\Gamma(\alpha)} + \frac{1}{2\Gamma(2+\alpha)}xT^H + o(T^H))$$

and we see that the at-the-money and skew terms are essentially the same as in the large deviations regime in (19). Since the answer is not surprising, we omit the details of the proof. To make this rigorous would require very fiddly tail estimates with Fourier arguments as in [EFGR19], which is beyond the scope of this article.

2.4 Fourier inversion formula for VIX calls for $T > 0$

Note we have the following Fourier inversion formula for exact pricing of VIX call options, where we have used Cauchy and Fubini's theorem in the first and second lines respectively:

$$\begin{aligned}
\mathbb{E}((\mathrm{VIX}_T-K)^+) &= \frac{1}{2\pi}\int_0^\infty (v^{\frac{1}{2}}-K)^+ \int_{-\infty}^\infty e^{-i(u-ia)v}\phi(u-ia,T)du dv \\
&= \frac{1}{2\pi}\int_{-\infty}^\infty \phi(u-ia,T)\int_0^\infty (v^{\frac{1}{2}}-K)^+ e^{-i(u-ia)v} dv du \\
&= \frac{1}{2\sqrt{\pi}}\int_{-\infty}^\infty \phi(u-ia,T)\frac{\mathrm{Erfc}(K\sqrt{a+iu})}{2(a+iu)^{\frac{3}{2}}}du
\end{aligned}$$

where $\phi(u,T):=e^{V_0 I^{1-\alpha}\psi_2(iu,T)}$ is the characteristic function of VIX_T^2, Erfc is the complementary error function and $a>0$ such that ia is inside the strip of analyticity of $\phi(.,T)$ (the condition $T_{\psi_2}(\mathrm{Re}(a))>T$ is sufficient for this, by the same reasoning as in the proof of Theorem 7 of [GGP19]).

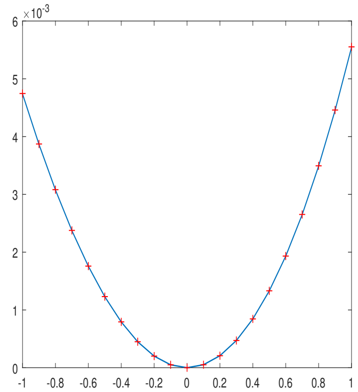


Figure 2: Here we have plotted $\bar{\Lambda}^{\rho=1}(pc)$ in blue using an Adams scheme with 2000 time steps applied to the driftless rough Heston VIE in Eq 24 in [FGS21] with $\rho = 1$ versus $T^{2H} \log \mathbb{E}(e^{\frac{p}{T^\alpha}(\mathrm{VIX}_T^2 - V_0)})$ (red) obtained using Monte Carlo for $T = .0001$, $\xi_0(t) = V_0 = .04$, $H = .25$, $\nu = .25$, and $\Delta = 1/12$ with 100,000 simulations, and we see both quantities are in close agreement.

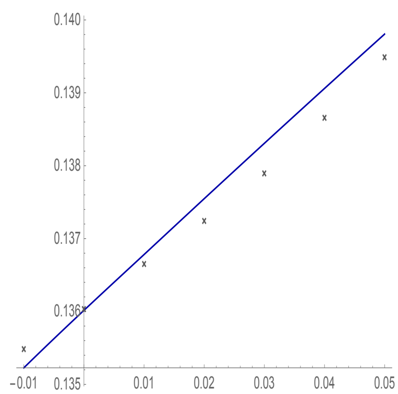


Figure 3: Here we have computed $\hat{\sigma}(x)$ using the same method as for Figure 3 in [FGS21] with 15 terms (blue) verses the VIX implied volatility computed using Monte Carlo (crosses) for $T = .0001$, $V_0 = 1$, $H = .25$, $\nu = .25$ and $\Delta = 1$ with 10,000,000 simulations and 200 time steps. It is difficult to verify exact agreement here since we can no longer exploit the usual Romano-Touzi/Willard conditioning trick for the Monte Carlo since ρ is effectively 1 here, and because of this the MC results for the left portion of the smile are less accurate since there is a significantly lower exponentially small probability that these (put) options expire in-the-money.

3 Ergodic behaviour of the Rough Heston model

In this section we return to the standard rough Heston model

$$V_t = V_0 + \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1}(\lambda(\theta - V_s)ds + \nu\sqrt{V_s}dW_s)$$

with non-zero mean reversion, i.e. $\lambda > 0$. Then from Theorem 4.3 in [ALP19], we know that

$$\mathbb{E}(e^{uV_T}) = e^{uV_0 + \lambda(\theta - V_0)\int_0^T \psi(s,u)ds + \frac{1}{2}V_0\nu^2 \int_0^T \psi(s,u)^2 ds} \tag{20}$$

for $\mathrm{Re}(u) \in [0,1]$, where

$$\psi = uK + (-\lambda\psi + \frac{1}{2}\nu^2\psi^2) * K = uK + I^\alpha(-\lambda\psi + \frac{1}{2}\nu^2\psi^2) \tag{21}$$

and $K(t) = \frac{1}{\Gamma(\alpha)}t^{\alpha-1}$, so $f = I^1\psi$ satisfies

$$f = uI^1K + I^{\alpha+1}(-\lambda\psi + \frac{1}{2}\nu^2\psi^2) = ut^{2\alpha-1}\frac{\Gamma(\alpha)}{\Gamma(\alpha)} - I^\alpha\lambda f + \frac{1}{2}\nu^2 I^{\alpha+1}\psi^2 .$$

The VIE in (21) is approximately linear for $t \gg 1$ since $\psi(t) \to 0$ as $t \to \infty$ because $K(t) \to 0$, so ignoring the quadratic term in (21) and taking the Laplace transform of both sides, we see that

$$\hat{\psi} = uz^{-\frac{1}{2}-H} - \lambda\hat{\psi}z^{-\frac{1}{2}-H}$$

so $\hat{\psi}(z) = \frac{u}{\lambda}\frac{\lambda}{z^\alpha+\lambda}$ which is the Laplace transform of $\psi^\infty(t) := ut^{\alpha-1}E_{\alpha,\alpha}(-\lambda t^\alpha) = \frac{u}{\lambda}\lambda t^{\alpha-1}E_{\alpha,\alpha}(-\lambda t^\alpha)$, where $E_{\alpha,\beta}(z)$ denotes the Mittag-Leffler function. From Appendix A in [ER19], we know that

$$\psi^\infty(t) \sim \frac{u}{\lambda}\frac{\alpha}{\lambda\Gamma(1-\alpha)}t^{-(\alpha+1)}$$

as $t \to \infty$, so ψ^∞ and $(\hat{\psi}^\infty)^2$ are clearly integrable at ∞. Hence from Lévy's convergence theorem, we expect that V_T tends weakly to a random variable V_∞ as $T \to \infty$ with characteristic function

$$\mathbb{E}(e^{iaV_\infty}) = e^{iaV_0 + \lambda(\theta - V_0)\int_0^\infty \psi(s,ia)ds + \frac{1}{2}V_0\nu^2\int_0^\infty \psi(s,ia)^2 ds}$$

for $a \in \mathbb{R}$. This may have implications for pricing European options if the V process evolves on a fast time scale, i.e.

$$V_t^{(\varepsilon)} = V_0^{(\varepsilon)} + \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\lambda\varepsilon^{-\alpha}(\theta - V_s^{(\varepsilon)})ds + \nu\varepsilon^{-H}\sqrt{V_s^{(\varepsilon)}}dW_s)$$

for $\varepsilon \ll 1$ (see Appendix B in [FGS21] to see why the powers of ε take this form), and we expect the leading order price of a European option with strike K and maturity T to be $C^{\mathrm{BS}}(S_0, K, (\int_0^\infty v\Pi(dv))^{\frac{1}{2}}, T)$ as $\varepsilon \to 0$, where $\Pi(dv)$ is the law of V_∞, and $C^{\mathrm{BS}}(S,K,\sigma,T)$ is the usual Black-Scholes formula for a call option with zero interest rates (see e.g. Fouque et al.[FPSS11] for background on singular perturbation theory for the $H = \frac{1}{2}$ case).

As a sanity check, for the case $H = \frac{1}{2}$, $K(t) = 1$ in (21), so the right hand side of (21) is just the term multiplying V_0 in the exponent in (20), but since $\psi(t) \to 0$ as $t \to \infty$, the left hand side (and hence) also the right hand side of (21) tends to zero. Hence we are just left with

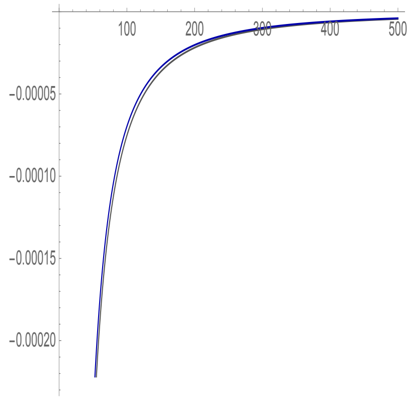
$$\mathbb{E}(e^{iaV_\infty}) = e^{\lambda\theta\int_0^\infty \psi(s,ia)ds}$$

i.e. the law of V_∞ is independent of V_0 for $H = \frac{1}{2}$ as we already know (it is an open question whether this remains true for $H < \frac{1}{2}$). For $H = \frac{1}{2}$, (21) is just an ODE, which we can solve explicitly to get $\psi(t,u) = \frac{2u\lambda}{u\nu^2 - e^{t\lambda}(u\nu^2 - 2\lambda)}$, from which we find that

$$\mathbb{E}(e^{uV_\infty}) = e^{\frac{\theta\lambda}{\nu^2}(\log 4 + 2\log\frac{\lambda}{2\lambda - u\nu^2})} = 4^{\frac{\theta\lambda}{\nu^2}}(\frac{\lambda}{2\lambda - u\nu^2})^{\frac{2\theta\lambda}{\nu^2}}$$

which is the mgf of a Gamma random variable with density $\Pi(x) = \frac{4^{\frac{\theta\lambda}{\nu^2}}e^{-\frac{2x\lambda}{\nu^2}}(\frac{x\lambda}{\nu^2})^{\frac{2\theta\lambda}{\nu^2}}}{x\Gamma(\frac{2\theta\lambda}{\nu^2})}$ for $x > 0$ (this result is of course well known).

Remark 3.1 For a log-normal rough Bergomi-type model where $V_t = V_0 e^{\int_0^t \kappa(t-s)dW_s}$ with Gamma kernel $\kappa(t) = \nu t^{H-\frac{1}{2}}e^{-\lambda t}$, we can easily verify that $\log V_\infty \sim N(0, \sigma_\infty^2)$, where $\sigma_\infty^2 = \lim_{t\to\infty}\int_0^t \kappa(t-s)^2 ds = \int_0^\infty \kappa(s)^2 ds = 4^{-H}\lambda^{-2H}\Gamma(2H)$, so V_∞ is log-normal.



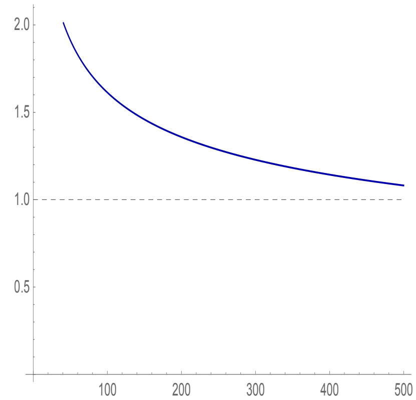


Figure 4: Here we have plotted the numerical solution to the VIE (21) using an Adams scheme (grey) verses the linearized approximation $\hat{\psi}^{\infty}$ (blue) for $\nu = .25$, $\lambda = 1$ and $H = .25$, and in the second graph we have plotted the ratio of these two terms.

References

[AGM18] Elisa Alòs, D.García-Lorite, A.Muguruza, "On smile properties of volatility derivatives and exotic products: understanding the VIX skew", preprint

[AE19] Abi Jaber, E. and O.El Euch, "Markovian structure of the Volterra Heston model", *Statistics&Probability Letters*, Volume 149, June 2019, Pages 63-72.

[ALP19] Abi Jaber, E., M.Larsson, and S.Pulido, "Affine Volterra processes", *Annals of Applied Probability*, Volume 29, Number 5 (2019), 3155-3200.

[ACLP19] Abi Jaber, C.Cuchiero, E., M.Larsson, and S.Pulido, "A weak solution theory for stochastic Volterra equations of convolution type", 2019, to appear in *Annals of Applied Probability.*

[Bil99] Billingsley, P., "Convergence of probability measures", Wiley, second edition, 1999.

[Bour20] Bourgey, F., "Stochastic approximations for financial risk computations", PhD thesis, L'Ecole Polytechnique, 2020.

[Brun17] Brunner, H., "Volterra integral equations", Cambridge University Press, Cambridge, 2017.

[CMS76] Chambers, J., C.Mallows, and B.Stuck, "A method for simulating stable random variables", 71 (1976), pp. 340-344.

[CPR17] Chavez, H.Prado and E.G.Reyes, "A Laplace transform approach to linear equations with infinitely many derivatives and zeta-nonlocal field equations", preprint, 2017

[CK13] Cont, R. and T.Kokholm, "A Consistent Pricing Model for Index Options and Volatility Derivatives", *Mathematical Finance*, Volume23, Issue 2, April 2013, Pages 248-27

[DJR19] Dandapani, A., P.Jusselin and M.Rosenbaum, "From quadratic Hawkes processes to super-Heston rough volatility models with Zumbach effect", to appear in *Quantitative Finance.*

[DZ98] Dembo, A. and O.Zeitouni, "Large deviations techniques and applications", Jones and Bartlet publishers, Boston, 1998.

[DeM18] De Marco, S., "Volatility derivatives in (rough) forward variance models", presentation, May 2018.

[EFGR19] El Euch, O., Fukasawa, M., Gatheral, J. and M.Rosenbaum, "Short-term at-the-money asymptotics under stochastic volatility models", *SIAM J. Finan. Math.*, 10(2), 491–511, 2019.

[EFR18] El Euch, O., M.Fukasawa, and M.Rosenbaum, "The microstructural foundations of leverage effect and rough volatility", *Finance&Stochastics*, 12 (6), p. 241-280, 2018.

[EGR18] El Euch, O., Gatheral, J. and M.Rosenbaum, "Roughening Heston", *Risk* magazine, pp. 84-89, May 2019.

[ER18] El Euch, O. and M.Rosenbaum, "Perfect hedging in Rough Heston models", *Annals of Applied Probability*, 28 (6), 3813-3856, 2018.

[EGLO21] Eckstein, S., G.Guo, T.Lim and J.Obloj, "Robust Pricing And Hedging Of Options On Multiple Assets And Its Numerics", *SIAM J. Finan. Math.*, 12(1), 158–188, 2021.

[ER19] El Euch, O. and M.Rosenbaum, "The characteristic function of Rough Heston models", *Mathematical Finance*, 29(1), 3-38, 2019.

[FS21] Forde, M., and B.Smith, "Rough Heston with jumps - joint calibration to SPX/VIX level and skew as $T \to 0$, and issues with the quadratic rough Heston model", preprint

[FGS21] Forde, M., S.Gerhold and B.Smith, "Small-time, large-time and $H \to 0$ asymptotics for the Rough Heston model", *Mathematical Finance*, 31(1), pgs 203-241, 2021.

[FSV21] Forde, M., B.Smith and L.Viitasaari, "Rough volatility and CGMY jumps with a finite history and the Rough Heston model - small-time asymptotics in the $k\sqrt{t}$ regime", with B.Smith and L.Viitasaari, *Quantitative Finance*, 21(4), 541-563, 21(4), 2021.

[FPSS11] Fouque, J.P., PAPANICOLAOU, G., Sircar, R. and Solna, K., "Multiscale Stochastic Volatility for Equity, Interest Rate and Credit Derivatives", Cambridge press, 2011

[Gath21] Gatheral, J., "Diamond trees and the forest expansion", Bloomberg Quant (BBQ) Seminar, January 2021 (joint work with E.Alòs, P.Friz and R.Radoičič).

[GR19] Gatheral, J. and R.Radoičič, "Rational Approximation of the Rough Heston Solution" *International Journal of Theoretical and Applied Finance*, Vol. 22, No. 3, 1950010, 2019

[GJR20] Gatheral, J., P.Jusselin and M.Rosenbaum, "The quadratic rough Heston model and the joint S&P 500/VIX smile calibration problem", RISK magazine, 2020.

[GK19] Gatheral, J. and M.Keller-Ressel, "Affine forward variance models", *Finance and Stochastics*, volume 23, pages 501-533, 2019.

[GGP19] Gerhold, S., C.Gerstenecker and A.Pinter, "Moment Explosions in the Rough Heston model", *Decisions in Economics and Finance*, 42(2), pp. 575-608, 2019.

[GLS90] Gripenberg, G, S.O.Londen, and O.Staffans, "Volterra Integral and Functional Equations", Cambridge University Press, 1990.

[Guy20] Guyon, J., "The Joint S&P 500/VIX Smile Calibration Puzzle Solved", *RISK* magazine, 2020.

[Guy20b] Guyon, J., "VIX-constrained Schrödinger bridges: joint S&P 500/VIX Smile Calibration calibration with continuous Stochastic Volatility Models", presentation, Nov 2020.

[Guy21] Guyon, J., "Dispersion-Constrained Martingale Schrödinger Problems and the exact joint S&P 500/Vix Smile Calibration puzzle", preprint, 2021

[GMN17] Guyon, J., R.Menegaux and M.Nutz, "Bounds for VIX futures given S&P 500 smiles", *Finance and Stochastics*, Volume 21, pages 593-630, 2017.

[HJT20] Horvath,B., A.Jacquier and P. Tankov, "Volatility options in rough volatility models", *SIAM Journal on Financial Mathematics*, 11(2): 437-469, 2020.

[HL19] Henry-Labordére, P., "From (Martingale) Schrodinger bridges to a new class of Stochastic Volatility Models', preprint, 2019

[JMP20] Jacquier, A. and A.Pannier, "Large and Moderate Deviations for Stochastic Volterra Systems", preprint

[JMP21] Jacquier, A., Muguruza, A. and A.Pannier, "Rough Multifactor Volatility for SPX and VIX options', preprint, 2021

[JR16] Jaisson, T. and M.Rosenbaum, "Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes", *Annals of Applied Probability*, 26(5), 2860-2882, 2016.

[JS13] Jacod, J. and A.Shiryaev, "Limit theorems for stochastic processes", volume 288, Springer Science&Business Media, 2013.

[Zha08] Zhang, X., "Euler schemes and large deviations for stochastic Volterra equations with singular kernels", *Journal of Differential Equations*, 244:2226–2250, 2008

[Vil09] Villani, C.: "Optimal transport. Old and new", volume 338 of Grundlehren der mathematischen Wissenschaften, Springer, 2009.

A Uniqueness of solutions to fractional Riccati VIEs

Following Theorem 3.1.2 and 3.1.4 in Chapter 3 in [Brun17], we consider a general non-linear VIE of the form

$$u(t) \;=\; \int_0^t \frac{1}{\Gamma(\alpha)}(t-s)^{H-\frac{1}{2}}(\frac{1}{2}(p+h(s))^2+\nu(p+h(s))u(s)+\frac{1}{2}\nu^2u(s)^2)ds \qquad \text{(A-1)}$$

where h is bounded and continuous, and suppose we have two continuous solutions u and u_2 to (A-1) on some interval $[0,T]$. Then

$$|u_2(t)-u(t)| \;\leq\; \int_0^t (t-s)^{H-\frac{1}{2}}L|u_2(s)-u(s)|ds$$

for some local Lipschitz constant L (since the function $\frac{1}{2}(p+h(s))^2+\nu(p+h(s))u+\frac{1}{2}\nu^2u^2$ is locally Lipschitz in u), and we write this more succinctly as $\Delta \leq -k*\Delta$, where $k(t) := -Lt^{H-\frac{1}{2}}$ and $\Delta = |u_2-u_1|$. The Laplace transform of k is $\hat{k}(\lambda) = -c\lambda^{-\alpha}$ where $c = L\Gamma(\alpha)$, and (from the definition of Eq 2.11 in [ALP19]) the resolvent r of k satisfies

$$k*r \;=\; k-r$$

which implies that

$$\hat{k}\hat{r} \;=\; \hat{k}-\hat{r}$$

and hence

$$\hat{r}(\lambda) \;=\; 1-(1+\hat{k}(\lambda))^{-1} \;=\; \frac{c}{c-\lambda^{\alpha}}\,.$$

Then $\hat{r}$ is the Laplace transform of $r(t) = -ct^{\alpha-1}E_{\alpha,\alpha}(ct^{\alpha})$ which is non-positive (see e.g. Table 1 in [ALP19] with $c \mapsto -c$ and end of proof of Proposition 2.1 in [FGS21]). Then using the following Lemma (taken from Appendix A.2 in [ACLP19]), we see that in fact $\Delta \equiv 0$, so we have uniqueness.

Lemma A.1 *(See Appendix A.2 in [ACLP19]. Suppose $f,g,k \in L^1([0,T])$. Assume k has non-positive resolvent r. Then if $f \leq g-k*f$, then $f \leq g-r*g$.*

Proof. Write $f+k*f = g-h$ for $h \geq 0$, so $\hat{f}+\hat{k}\hat{f} = \hat{g}-\hat{h}$. Then from the definition of the resolvent: $\hat{k}\hat{r} = \hat{k}-\hat{r}$ we find that

$$\begin{aligned} \hat{f}+\frac{\hat{r}}{1-\hat{r}}\hat{f} &= \hat{g}-\hat{h} \\ \Rightarrow \quad \hat{f}(1-\hat{r})+\hat{r}\hat{f} = \hat{f} &= \hat{g}-\hat{h}-\hat{r}(\hat{g}-\hat{h}) \end{aligned}$$

so $f = g-h-r*(g-h) \leq g-r*g$. ■

B Derivation of the VIE

Theorem B.1 *(minor variant of Theorem 7.1 in [ALP19] without their restriction that* $\mathrm{Re}(\psi_1) \in [0,1]$ *and no drift term). Consider the d-dimensional stochastic convolution equation:*

$$X_t \;=\; X^0(t)+\int_0^t \tilde{K}(t-s)\sigma(X_s)dW_s$$

so $\mathbb{E}(X_t) = X^0(t)$ for $\xi \in L^1$, where $\tilde{K} \in L^2([0,T];\mathbb{R}^{d\times d})$, $a(x) = \sigma^T(x)\sigma(x) = x_1A^1+...+x_dA^d$, A^i is a (symmetric) $d\times d$ matrix for each $i = 1..d$, $A(u) = (uA^1u^T, uA^2u^T, ..., uA^du^T)$, W is a d-dimensional Brownian motion, and

$$\psi \;=\; u\tilde{K}+(f+\frac{1}{2}A(\psi))*\tilde{K} \qquad \text{(B-1)}$$

for $f \in L^1$ and let Y satisfy $dY_t = -\frac{1}{2}\psi(T-t)^2\sigma(X_t)^2dt+\psi(T-t)\sigma(X_t)dW_t$ with

$$Y_0 \;=\; uX^0(T)+(f*X_0)_T+\frac{1}{2}\int_0^T \psi(T-s)a(X^0(t))\psi(T-s)^T ds\,.$$

Then if ψ is bounded on $[0,T]$, then e^Y is an $\mathcal{F}_t^W$-martingale on $[0,T]$ and we have the exponential-affine formula:

$$\mathbb{E}(e^{uX_T+(f*X)_T}|\mathcal{F}_t^W) \;=\; e^{Y_t}\,.$$

For our specific case of interest for the Rough Heston model, $f_2 = 0$, $u_2 = 0$ and $X^0(t) = (0, \xi_0(t))$ so we can re-write (B-1) *in component form as*

$$\begin{aligned}\psi_1 &= u_1 + f_1 * 1\\ \psi_2 &= u_2 K + \frac{1}{2}(\psi_1^2 + 2\nu\psi_1\psi_2 + \nu^2\psi_2^2) * K \quad = \quad I^\alpha F(\psi_1, \psi_2)\end{aligned}$$

where $K(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)}$ *and* $F(\psi_1, \psi_2) = \frac{1}{2}(\psi_1 + \nu\psi_2)^2$, *and* $Y_0 = uX^0(T) + (f * X_0)_T + \frac{1}{2}\int_0^T \psi(T - s)a(X^0(T))\psi(T - s)^T ds = I^1(\xi_0(T - (.))D^\alpha\psi_2)(T)$, *which further simplifies to the familiar expression* $V_0 I^{1-\alpha}\psi_2(T)$ *if* $\xi_0(.)$ *is flat.*

Proof. We let $\mathcal{F}_t := \mathcal{F}_t^W$ throughout, and we first note that

$$\mathbb{E}(X_s|\mathcal{F}_t) = X_0(s) + \int_0^{s\wedge t} \tilde{K}(s - v)dM_v \qquad \text{(B-2)}$$

where $dM_t = \sigma(X_t)dW_t$. Now let

$$Y_t = \mathbb{E}(uX_T + \int_0^T f(T - s)X_s ds|\mathcal{F}_t) + \frac{1}{2}(\int_0^T - \int_0^t)\psi(T - s)a(\mathbb{E}(X_s|\mathcal{F}_t))\psi(T - s)^T ds$$

for $t \le T$. Using (B-2) and the affine property of $a(.)$ we can re-write Y_t in the form $Y_0 + (...)$ as

$$\begin{aligned}Y_t &= uX^0(T) + (f * X_0)(T) + \frac{1}{2}\int_0^T \psi(T - s)a(X_0(s))\psi(T - s)^T ds\\ &+ u\int_0^t \tilde{K}(T - s)dM_s + \int_0^T f(T - s)\int_0^{s\wedge t} \tilde{K}(s - v)dM_v ds + \frac{1}{2}\int_0^T \psi(T - s)a(\int_0^{s\wedge t} \tilde{K}(s - v)dM_v)\psi(T - s)^T ds\,.\\ &- \frac{1}{2}\int_0^t \psi(T - s)a(X_s)\psi(T - s)^T ds\end{aligned}$$

(the sum of the three terms on the right hand side in the first line here is Y_0). From Fubini we see that the fifth term here can be re-written as

$$\int_0^T f(T - s)\int_0^{s\wedge t} \tilde{K}(s - v)dM_v ds = \int_0^t\int_v^T f(T - s)\tilde{K}(s - v)ds dM_v\,.$$

Similarly

$$\begin{aligned}\int_0^T \psi(T - s)a(\int_0^{s\wedge t} \tilde{K}(s - v)dM_v)\psi(T - s)^T ds &= \int_0^T \psi(T - s)(\sum_{i=1}^d A^i\int_0^{s\wedge t} \tilde{K}(s - v)dM_v^i)\psi(T - s)^T ds\\ &= \int_0^t\int_v^T A(\psi(T - s))\tilde{K}(s - v)ds dM_v\end{aligned}$$

and recall that $A(u) = (uA^1(u)u^T, uA^2(u)u^T, ..., uA(u)^d u^T)$. Thus

$$\begin{aligned}Y_t &= Y_0 + u\int_0^t \tilde{K}(T - v)dM_v + \int_0^t\int_v^T f(T - s)\tilde{K}(s - v)ds dM_v + \frac{1}{2}\int_0^t\int_v^T A(\psi(T - s))\tilde{K}(s - v)ds dM_v\\ &- \frac{1}{2}\int_0^t \psi(T - s)a(X_s)\psi(T - s)^T ds\\ &= Y_0 + \int_0^t (u\tilde{K}(T - v) + \int_v^T f(T - s)\tilde{K}(s - v)ds + \frac{1}{2}\int_v^T A(\psi(T - s))\tilde{K}(s - v)ds)dM_v\\ &- \frac{1}{2}\int_0^t \psi(T - s)a(X_s)\psi(T - s)^T ds\end{aligned}$$

and we note that

$$\int_v^T f(T - s)\tilde{K}(s - v)ds = \int_0^{T-v} f(T - (s + v))\tilde{K}(s)ds = (\tilde{K} * f)(T - v)$$

and similarly $\displaystyle\int_v^T A(\psi(T - s))\tilde{K}(s - v)ds = (A(\psi) * \tilde{K})(T - v)$

so

$$Y_t = Y_0 + \int_0^t (u\tilde{K}(T - v) + (f * \tilde{K})(T - v) + \frac{1}{2}(A(\psi) * \tilde{K})(T - v))dM_v - \frac{1}{2}\int_0^t \psi(T - s)a(X_s)\psi(T - s)^T ds\,.$$

Comparing this expression to the "Driftless" Ricccati eq:

$$\psi \;=\; u\tilde{K} + (f + \frac{1}{2}A(\psi)) * \tilde{K}$$

we see that

$$Y_t \;=\; Y_0 + \int_0^t \psi(T-s)\sigma(X_s)dW_s - \frac{1}{2}\int_0^t \psi(T-s)a(X_s)\psi(T-s)^T ds$$

so e^{Y_t} is a local martingale. If e^Y is a true martingale on $[0,T]$ (see the end of the proof for clarification on this point), then $\mathbb{E}(e^{Y_T}|\mathcal{F}_t) = e^{Y_t}$ and in particular

$$\mathbb{E}(e^{Y_T}) \;=\; \mathbb{E}(e^{uX_T+(f*X)_T}) \;=\; e^{Y_0} \;=\; e^{uX^0(T)+(f*X_0)(T)+\frac{1}{2}\int_0^T \psi(T-s)a(X_0(s))\psi(T-s)^T ds}. \quad \text{(B-3)}$$

In our specific case $X_t = \begin{pmatrix}\tilde{X}_t \\ V_t\end{pmatrix}$ with kernel $\tilde{K} = \begin{pmatrix}1 & 0 \\ 0 & K\end{pmatrix}$ and $X^0(t) = (0, \xi_0(t))$. Then $\sigma(X_t) = \sqrt{V_t}\begin{pmatrix}0 & 1 \\ 0 & \nu\end{pmatrix}$ so

$$a(X_t) \;:=\; \sigma(X_t)\sigma^T(X_t) \;=\; V_t\begin{pmatrix}0 & 1 \\ 0 & \nu\end{pmatrix}\begin{pmatrix}0 & 0 \\ 1 & \nu\end{pmatrix} \;=\; V_t\begin{pmatrix}1 & \nu \\ \nu & \nu^2\end{pmatrix}$$

which implies that $A^1 = 0$ and

$$A^2(\psi) \;=\; \begin{pmatrix}\psi_1 & \psi_2\end{pmatrix}\begin{pmatrix}1 & \nu \\ \nu & \nu^2\end{pmatrix}\begin{pmatrix}\psi_1 \\ \psi_2\end{pmatrix} \;=\; \psi_1^2 + 2\nu\psi_2 + \nu^2\psi_2^2\,.$$

Then the Riccati-Volterra eq becomes:

$$\begin{aligned}\psi \;=\; \begin{pmatrix}\psi_1 & \psi_2\end{pmatrix} \;&=\; u\tilde{K} + (f + \frac{1}{2}A(\psi)) * \tilde{K} \\ &=\; (u_1, u_2)\begin{pmatrix}1 & 0 \\ 0 & K\end{pmatrix} + ((f_1, 0) + \frac{1}{2}(0, \psi_1^2 + 2\nu\psi_1\psi_2 + \nu^2\psi_2^2)) * \begin{pmatrix}1 & 0 \\ 0 & K\end{pmatrix}\end{aligned}$$

which we can re-write as

$$\begin{aligned}\psi_1 \;&=\; u_1 + f_1 * 1 \\ \psi_2 \;&=\; u_2 K + \frac{1}{2}(\psi_1^2 + 2\nu\psi_1\psi_2 + \nu^2\psi_2^2) * K \;=\; u_2 K + I^\alpha F(\psi_1, \psi_2)\,. \qquad \text{(B-4)}\end{aligned}$$

and

$$Y_0 \;=\; u_2\xi_0(T) + (f * X)_0 + \frac{1}{2}\int_0^T \psi(T-s)a(X_0(s))\psi(T-s)^T ds$$

and

$$\frac{1}{2}\psi(T-s)a(X^0(t))\psi^T(T-s) \;=\; \xi_0(t)(\psi_1(T-s), \psi_2(T-s))\begin{pmatrix}1 & \nu \\ \nu & \nu^2\end{pmatrix}\begin{pmatrix}\psi_1(T-s) \\ \psi_2(T-s)\end{pmatrix} \;=\; \xi_0(t)(\psi_1^2 + 2\nu\psi_1\psi_2 + \nu^2\psi_2^2)$$

so

$$\begin{aligned}Y_0 \;&=\; u_2\xi_0(T) + (f * X)_0 + \frac{1}{2}\int_0^T \psi(s)a(X_0(T-s))\psi(s)^T ds \\ &=\; u_2\xi_0(T) + (f * X)_0 + \frac{1}{2}\int_0^T \xi_0(T-s)(\psi_1(s)^2 + 2\nu\psi_1(s)\psi_2(s) + \nu^2\psi_2(s)^2)ds \\ &=\; u_2\xi_0(T) + (f * X)_0 + I^1(\xi(T-(.))D^\alpha(\psi_2 - u_2 K))(T) \\ &=\; (f * X)_0 + I^1(\xi(T-(.))D^\alpha\psi_2)(T)\end{aligned}$$

(where we have used (B-4) for the second equality and $(I^{1-\alpha}K)(t) = 1$ for the final equality.) which is the exponent in (B-3). Moreover

$$Y_t \;=\; \xi_0(T) + \int_0^t \psi(T-s)\sigma(X_s)dW_s - \frac{1}{2}\int_0^t \psi(T-s)a(X_s)\psi(T-s)^T ds$$

so

$$\begin{aligned}dY_t \;&=\; \psi(T-t)\sigma(X_t)dW_t - \frac{1}{2}\psi(T-t)a(X_t)\psi(T-t)^T dt \\ &=\; \sqrt{V_t}(\psi_1(T-t) + \nu\psi_2(T-t))dW_t^2 - \frac{1}{2}V_t(\psi_1(T-t)^2 + \nu\psi_1(T-t))^2 dt\,.\end{aligned}$$

Then from Lemma 7.3 in [ALP19], e^Y is a genuine $\mathcal{F}_t^W$-martingale on $[0,T]$ if $\psi_1 + \nu\psi_2 \in L^\infty([0,T])$ and since f is integrable, $\psi_2 \in L^\infty$ implies $\psi_1 + \nu\psi_2 \in L^\infty$, and $\psi_2 \in L^\infty$ if $T^*(u) > T$ (where $T^*(u)$ is the explosion time for ψ_2) since the solution to the VIE for ψ_2 is continuous up to the explosion time.
■

C Uniform moment bound

Lemma C.1 *(see also Lemma 3.1 in [ALP19] and Lemma A.1 in [JMP20]). For $m \geq 2$*

$$\sup_{t \leq T} \mathbb{E}(V_t^m) \quad \leq \quad c_{m,T}$$

for some finite constant $c_{m,T}$ which depends on m and T and the model parameters.

Proof. Setting $K(t) = t^{\alpha-1}/\Gamma(\alpha)$ and using the Bukholder-Davis-Gundy inequality applied to the martingale $M_u := \int_0^u K(t-s)\sqrt{V_s}dW_s$ at $t = u$, we see that

$$\begin{aligned}
\mathbb{E}(V_t^m) &= \mathbb{E}((\xi_0(t) + \int_0^t K(t-s)\sqrt{V_s}dW_s)^m) \leq 2^m \xi_0(t)^m + 2^m C_m \mathbb{E}((\int_0^t K(t-s)^2 V_s ds)^{\frac{1}{2}m}) \\
&= 2^m \xi_0(t)^m + 2^m C_m \mathbb{E}((\int_0^t K(t-s)^{2-\frac{4}{m}} K(t-s)^{\frac{4}{m}} V_s ds)^{\frac{1}{2}m}) \\
&\leq 2^m \xi_0(t)^m + 2^m C_m \|K\|_2^{m-2} \int_0^t K(t-s)^2 \mathbb{E}(V_s^{\frac{1}{2}m}) ds \\
&\leq 2^m \xi_0(t)^m + 2^m C_m \|K\|_2^{m-2} \int_0^t K(t-s)^2 \mathbb{E}(a(1+V_s)^m) ds
\end{aligned}$$

where we have used Hölder's inequality with $p = \frac{1}{2}m$ and $q = (1-1/p)^{-1}$ in the final line as in Appendix A.2 in [JMP20], so $f(t) := \mathbb{E}(V_t^m)$ satisfies

$$f(t) \leq c + c\int_0^t K(t-s)^2 f(s)ds = c + c\int_0^t (t-s)^{\alpha_2 - 1} f(s)ds$$

for some constant $c > 0$ and $t \in [0,T]$, where $\alpha_2 = 2H$. Using Lemma A.1, we know that

$$\begin{aligned}
f(t) &\leq c - (r * c)(t) \\
&= c + c\int_0^t (t-s)^{\alpha_2 - 1} E_{\alpha_2,\alpha_2}(\tilde{c}(t-s)^{\alpha}) c ds < \infty
\end{aligned}$$

where r is the resolvent of ct^{α_2-1} given by $\hat{r}(t) = -\tilde{c}t^{\alpha_2-1}E_{\alpha_2,\alpha_2}(\tilde{c}t^{\alpha_2})$ where $\tilde{c} = c\Gamma(\alpha_2)$. $E_{\alpha_2,\alpha_2}(\tilde{c}(t-s)^{\alpha_2})$ is bounded on $[0,t]$, so $f(t) \leq const. \times \int_0^t (t-s)^{\alpha_2-1}\tilde{c}s^{-\alpha_2}ds < \infty$ for all $s \in [0,t]$. ■

D Asymptotics for VIX call options

From Jensen's inequality, we know that for any $q \geq 1$ we have

$$(\mathrm{VIX}_T^2)^q = (\frac{1}{\Delta}\int_T^{T+\Delta} \xi_T(u)du)^q \leq \frac{1}{\Delta}\int_T^{T+\Delta} \xi_T(u)^q du$$

and hence

$$\begin{aligned}
\mathbb{E}(\mathrm{VIX}_T^{2q}) &\leq \frac{1}{\Delta}\int_T^{T+\Delta} \mathbb{E}(\xi_T(u)^q)du = \frac{1}{\Delta}\int_T^{T+\Delta} \mathbb{E}(\,\mathbb{E}(V_u|\mathcal{F}_T)^q)du \\
&\leq \frac{1}{\Delta}\int_T^{T+\Delta} \mathbb{E}(V_u^q)du \qquad \text{(D-1)}
\end{aligned}$$

which will be needed further down.

- Lower bound. We first note that for x fixed and any $\delta \in (0,x)$, $e^{xT^{\frac{1}{2}-H}} \leq 1 + (x+\delta)T^{\frac{1}{2}-H}$ for T sufficiently small. Recall that $\mathrm{VIX}_0^2 = \frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(u)du$ and we set $k_{x,\delta} := \mathrm{VIX}_0(x+\delta)$. We first note that for $\delta > 0$ and $T = T(\delta)$ sufficiently small, $e^{xT^{\frac{1}{2}-H}} \leq 1 + (x+\delta)xT^{\frac{1}{2}-H}$. Thus for $T = T(\delta)$ sufficiently small

$$\begin{aligned}
\mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0 e^{xT^{\frac{1}{2}-H}})_+) &\geq \mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0(1 + (x+\delta)T^{\frac{1}{2}-H})_+) \\
&= T^{\frac{1}{2}-H}\mathbb{E}((\frac{\mathrm{VIX}_T - \mathrm{VIX}_0}{T^{\frac{1}{2}-H}} - k_{x,\delta})_+) \\
&\geq \delta T^{\frac{1}{2}-H}\mathbb{E}(1_{\frac{\mathrm{VIX}_T - \mathrm{VIX}_0}{T^{\frac{1}{2}-H}} > k_{x,\delta}+\delta}) \\
&= \delta T^{\frac{1}{2}-H}\mathbb{P}(\mathrm{VIX}_T > \mathrm{VIX}_0 + T^{\frac{1}{2}-H}(k_{x,\delta}+\delta)) \\
&= \delta T^{\frac{1}{2}-H}\,\mathbb{P}(\mathrm{VIX}_T^2 > \mathrm{VIX}_0^2 + 2\mathrm{VIX}_0\,(k_{x,\delta}+\delta)T^{\frac{1}{2}-H} + (k_{x,\delta}+\delta)^2 T^{1-2H}).
\end{aligned}$$

But for $T = T(\delta)$ sufficiently small, the right hand side here is greater than or equal to

$$\begin{aligned} & \delta T^{\frac{1}{2}-H}\,\mathbb{P}(\mathrm{VIX}_T^2 - \mathrm{VIX}_0^2 > 2\mathrm{VIX}_0\,(k_{x,\delta}+2\delta)T^{\frac{1}{2}-H}) \\ = \;& \delta T^{\frac{1}{2}-H}\,\mathbb{P}(\frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(u)du - \frac{1}{\Delta}\int_T^{T+\Delta}\xi_0(u)du > 2\mathrm{VIX}_0\,(k_{x,\delta}+2\delta)T^{\frac{1}{2}-H})\,. \end{aligned}$$

Then using the LDP and the continuity of J we see that

$$\liminf_{T\to 0} T^{2H}\log\mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0\,e^{xT^{\frac{1}{2}-H}})_+) \;\;\geq\;\; -J(2\mathrm{VIX}_0(k_{x,\delta}+2\delta)) \;\;=\;\; -J(2\mathrm{VIX}_0^2 + 2\delta\mathrm{VIX}_0 + 4\delta\mathrm{VIX}_0))\,.$$

We then let $\delta \to 0$ and again use the continuity of the rate function $J(x)$ to obtain the required lower bound.

- Upper bound. From Hölder's inequality, we note that for $q > 1$

$$\begin{aligned} \mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0\,e^{xT^{\frac{1}{2}-H}})_+) \;&\leq\; \mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0(1+xT^{\frac{1}{2}-H})))_+ \\ &=\; \mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0(1+xT^{\frac{1}{2}-H}))_+ 1_{\mathrm{VIX}_T\geq \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H})}) \\ &\leq\; \mathbb{E}[(\mathrm{VIX}_T - \mathrm{VIX}_0(1+xT^{\frac{1}{2}-H}))_+^q]^{\frac{1}{q}}\,\mathbb{E}(1_{\mathrm{VIX}_T\geq \mathrm{VIX}_0\,+xT^{\frac{1}{2}-H}})^{1-\frac{1}{q}}\,. \end{aligned}$$

Thus

$$\begin{aligned} & T^{2H}\log\mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H}))_+) \\ \leq\;& \frac{T^{2H}}{q}\log\mathbb{E}[(\mathrm{VIX}_T - \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H}))_+^q] \;+\; T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T \geq \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H})) \\ \leq\;& \frac{T^{2H}}{q}\log\mathbb{E}(\mathrm{VIX}_T^q) \;+\; T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T \geq \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H})) \\ \leq\;& \frac{T^{2H}}{q}\log(\mathbb{E}(\mathrm{VIX}_T^{2q})^{\frac{1}{2}}) \;+\; T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T \geq \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H})) \\ \leq\;& \frac{T^{2H}}{q}\frac{1}{2}\log(\frac{1}{\Delta}\int_T^{T+\Delta}\mathbb{E}(V_u^q)du) \;+\; T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T^2 \geq \mathrm{VIX}_0^2(1+xT^{\frac{1}{2}-H})^2) \\ & \text{(by (D-1))} \\ \leq\;& \frac{T^{2H}}{q}\frac{1}{2}\log(\frac{1}{\Delta}\int_T^{T+\Delta}(\mathbb{E}(V_u^q)^{\frac{1}{q}})^q du \;+\; T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T^2 \geq \mathrm{VIX}_0^2(1+2xT^{\frac{1}{2}-H})) \\ & \text{(using Minkowski applied to } \mathbb{E}((V_u)^q)) \\ \leq\;& \frac{T^{2H}}{q}\frac{1}{2}\log(c_{q,T}^{\frac{1}{q}})^q \;+\; T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T^2 \geq \mathrm{VIX}_0^2(1+2xT^{\frac{1}{2}-H}) \end{aligned}$$

for some finite constant $c_{q,T}$ depending on q and T, where we have used Lemma C.1 in the final line. Letting $T \to 0$ in the final line and using the LDP and the continuity of J, and then letting $q \to \infty$, we see that

$$\limsup_{T\to 0} T^{2H}\log\mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H}))_+) \;\;\leq\;\; -J(2\mathrm{VIX}_0^2 x)\,.$$