Rough Heston with jumps - joint calibration to SPX/VIX level and skew as $T \to 0$, and issues with the quadratic rough Heston model

Martin Forde Benjamin Smith*

29th March 2024

Abstract

We augment the well known rough Heston model with an additional independent CGMY/KoboL-type jump process with a Brownian component, and 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 this sense, the additional CGMY component allows the SPX and VIX smiles to decouple as $T \to 0$ in some sense (which is not possible for a rough model without jumps), and the model can generate power-law skew for both smiles with different effective H-values. We also formally compute small-time Edgeworth asymptotics for implied volatility for the quadratic rough Heston model of Gatheral et al.[GJR20] and we find that the short-time skew is positive if $Z_0 > b$ which also means the short-time skew can flip sign as time evolves (since the history will not be relevant in the small-time limit) and we discuss some additional properties/drawbacks of this model. [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 apparently 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 (this issue will be partially addressed in Section 3 of this article).

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 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

*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

formula with the VIX future price $\mathbb{E}(\text{VIX}_T)$, 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 are still comparatively high), and empirical and model-generated VIX implied volatility smiles have always exhibited a marked positive skew (see plots in [GJR20],[Guy20],[DeM18],[HJT20] et al.).

In [FGS21b], we consider a generalized Rough Heston model with non-zero initial variance curve $\xi_0(t)$, and we show that (module a scaling factor) $(\text{VIX}_T^2 - \text{VIX}_0^2)/T^{\frac{1}{2}-H}$ satisfies the same small-time LDP as as the re-scaled log stock price $(X_T - X_0)/T^{\frac{1}{2}-H}$ in the main Theorem 3.3 in [FGS21] if we set $\rho = 1$. 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.

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, we enrich the model in this article with an additional independent CGMY/KoBoL process L (c.f. [BL00]) as in [FSV21] with $Y \in (1,2)$ (which implies L has infinite variation), 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 the SPX and VIX smiles decouple in some sense as $T \to 0$ if $1 - Y/2 < H$.

More specifically, we can use the rough Heston parameters H, V_0, ν and the initial value VIX_0 of the VIX to fit the overall level and the at-the-money skew of VIX options as $T \to 0$. We can then use the Brownian component σ and the CGMY jump parameters Y, C_+ and C_- to fit the observed at-the-money implied volatility as $T \to 0$, the first order at-the-money correction and the at-the-money skew of SPX options as $T \to 0$. Finally we choose the CGMY parameters G and/or M and $\xi_0(t) = \mathbb{E}(V_t)$ in the definition of the V process to be consistent with the observed variance term structure and the value that we have just calibrated for V_0. The upshot of this is that our CGMY-rough Heston model can generate different power-law skew for the SPX and VIX smiles in the $T \to 0$ limit. [Guy20b] provides a list of references on earlier papers which have also attempted to decouple the SPX and VIX smiles using jumps (e.g. Cont&Kokhom[CK13] which uses a (non-rough) Bergomi-type model driven by a Brownian motion plus a Lévy component), but all were written in the pre-rough stoc vol era, and no exact results were found to the best of our knowledge.

In the final section, we formally compute small-time asymptotics for implied volatility for the quadratic rough Heston model of Gatheral et al.[GJR20] in the usual Edgeworth regime where the log-moneyness scales like $z\sqrt{T}$ for z fixed as $T \to 0$, and we find that the short-time skew is positive if $Z_0 > b$.

Given admissible marginals (μ_1, μ_V, μ_2) for $(S_1, V, S_2) := (S_{T_1}, \text{VIX}_{T_1}, S_{T_2})$, Guyon[Guy21] (see also [Guy20]) describes how to construct an admissible trivariate law for these three quantities (i.e. in the convex set $\mathcal{P}(\mu_1, \mu_V, \mu_2)$ using their notation) using a dual formulation of $\inf_{\mu \in \mathcal{P}(\mu_1, \mu_V, \mu_2)} H(\mu|\bar{\mu})$ where H denotes the relative entropy function (i.e. the Kullback-Leibler divergence) and $\bar{\mu}$ is a reference probability measure not in $\mathcal{P}(\mu_1, \mu_V, \mu_2)$. The proof of the main Theorem 16 in [Guy21] uses the Sion minimax theorem, which is justified since their $\Pi(\mu_1, \mu_V, \mu_2)$ set is weakly compact (cf. Lemma 4.4 in Villani[Vil09]) and $H(\mu|\bar{\mu})$ is weakly lower semicontinuous in μ (because this is the rate function for Sanov's theorem in large deviations theory), and some other semicontinuity properties of the quantities which appear in the dual formulation, see the end of pg 27 in [Guy21] for details. The argmin for the inner inf in the minmax problem is computed explicitly and the optimal "model" is given by a Gibbs-type trivariate probability measure for (S_1, V, S_2), and this dual problem is solved numerically using the Sinkhorn (exponentially fast) fixed point scheme by deriving the associated extremal equations for the optimal u_1, u_V, u_2 and Δ_S^S, Δ_L^L (by setting the Gateaux derivatives to zero in the usual way, see Eqs 6.1 and 6.2), see also [EGLO21] for further discussion on the Sinkhorn method. This approach only yields an admisible joint trivariate law for (S_1, V, S_2) (essentially a discrete-time model with two time-steps), so although it can be used for consistent pricing of forward-start options at time zero, it cannot directly be used to price other path-dependent contracts, and does not give us a dynamic continuous-time model with stylized features such as power-law skew, Hölder continuous sample paths for the volatility with $H < \frac{1}{2}$, Zumbach effect etc. Theorem 12 part iii) in [Guy21] gives an if and only convex ordering condition for the S_1, S_2 and V marginals to be arbitrage-free (this condition also implicitly assumes the centering conditions $\mathbb{E}(S_1) = \mathbb{E}(S_2)$ and $\mathbb{E}(L(S_2)) = \mathbb{E}(L(S_1)) + \mathbb{E}(V^2)$ are satisfied, where $L(x) = -\frac{2}{\tau}\log x$), but unfortunately this convex ordering condition is impossible to check in practice since no simple class of test functions exists in dimension more than 1 (see comment at the end of page 11 in [GMN17]), unless the trivariate law comes from a known model, in which case the [Guy21] calibration methodology is not required.

The related article [GMN17] prove the absence of a duality gap for the robust superhedging problem for a VIX future given tradeable SPX options at two maturities and forward-starting log contracts, and using a simple conditional Jensen argument show that the extremal price is attained by a model where the VIX at T_1 is just a deterministic function of S_{T_1}. [GMN17] also investigate so-called functionally generated portfolios for this problem which are tractable and optimal in certain cases and improve on the classical bounds, and they also approximate the true solution numerically via a linear programming problem using the Mosek software package.

Guo et al.[GLOW19] derive a novel method for constructing a local-stochastic volatility model consistent with a finite number of European options by optimizing a certain cost function and re-casting the problem as a convex optimisation problem, and numerically solving the dual problem, which involves a non-linear HJB equation, but although very interesting this approach has essentially the same type of issue as [Guy21], namely that there is no easy way to know if such a model exists to begin with.

Gatheral et al.[Gath21] introduce the diamond tree formalism and forest expansions which in particular yields a new exponential-affine formula for the triple mgf $\mathbb{E}(e^{aX_T+b\langle X\rangle_T+c\xi_T(T)})$ under the rough Heston model, which is a variation of Theorem 7.1 in [ALP19]. Lemma 2.6 in [ALP19] also contains an inversion formula for a general stochastic convolution equation. [Guy20] and [DeM18] consider a mixed one-factor rough Bergomi model with two different vol-of-vol parameters but driven by the same Brownian motion so $\xi_T(u)$ is now the sum of two lognormal random variables (and hence no longer lognormal), which allows for more right skew in fitting VIX smiles at non-zero maturities (the distribution of VIX_T under the standard rBergomi model is close to lognormal since Δ is small, so rBergomi is somewhat limited in producing skew for VIX smiles). [Bour20] computes small-Δ expansions for VIX call/put options, where for rough Bergomi the leading order term is lognormal since $\xi_T(u)$ is lognormal and the correction terms are expressed in terms of Black-Scholes Greeks of a call/put on the aforementioned lognormal random variable. We also mention [HJT20] who consider so-called "Modulated Volterra" models which are essentially rough Bergomi-type models but with an additional independent stochastic factor in the Volterra integral (similar to Brownian semi-stationary processes but without the stationarity). They focus on the specific case when the additional factor is a time-homogeneous positive conservative process independent of the other driving Brownian motions (which includes the CIR process a special case), and they also report greater flexibility in fitting VIX smiles using this class of models.

2 The model

Consider a generalized Rough Heston model for a log stock price process $X_t = \log S_t$:

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

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. 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 [JP20] 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]).

2.1 Adding jumps - decoupling the SPX and VIX short-maturity skews

We now augment our rough Heston model to include jumps of infinite variation. Specifically (following [FSV21]), we now assume that the log stock price process X evolves as

$$dX_t = -\frac{1}{2}V_t dt + \sqrt{V_t}(\rho dW_t + \bar{\rho} dB_t) - dL_t \tag{2}$$

where L is a Lévy process independent of W and B with $\mathbb{E}(e^{-uL_t}) = e^{tV(u)}$ for u in some interval (u_-, u_+) which includes 0 and 1 with $\mathbb{E}(e^{-L_t}) = 1$ (i.e. $V(1) = 0$) which ensures that L is a martingale with respect to its own filtration, and we assume that $X_0 = 0$ and V is not equal to the zero function W.L.O.G.

Then

$$-\frac{2}{\Delta}\mathbb{E}(X_{T+\Delta}-X_T|\mathcal{F}_T) \;=\; \frac{1}{\Delta}\mathbb{E}(\int_T^{T+\Delta} V_u|\mathcal{F}_T) \;-\; \frac{2}{\Delta}\mathbb{E}(-(L_{T+\Delta}-L_T)|\mathcal{F}_T)$$
$$= \frac{1}{\Delta}\int_T^{T+\Delta} \xi_T(u)du \;+\; a_1$$

where $\xi_t(u) := \mathbb{E}(V_u|\mathcal{F}_t)$ as before, and

$$a_1 \;:=\; -2V'(0) \;\geq\; -2\log\mathbb{E}(e^{-(L_{T+\Delta}-L_T)}|\mathcal{F}_T) \;=\; -2\log 1 \;=\; 0 \tag{3}$$

and we have used Jensen and the martingale condition to obtain the inequality here. Moreover, since $V(0)=V(1)=0$, $V'(0)<0$ (since we are assuming that V is not identically zero), so we see that a_1 is strictly positive.

Then we see that $\mathrm{VIX}_T^2 = \frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(u)du \,+\, a_1$, so the effect of the jumps is to increase the value of VIX_T^2 by a *shift factor*[2] equal to a_1. Thus $(\mathrm{VIX}_T^2 - a_1 - \frac{1}{\Delta}\int_0^{\Delta}\xi_0(u)du)/T^{\frac{1}{2}-H} = (\mathrm{VIX}_T^2-\mathrm{VIX}_0^2)/T^{\frac{1}{2}-H}$ now satisfies the LDP as $T\to 0$ with speed T^{-2H} and rate function $J(x)$ as before, and we have the following:

Corollary 2.1

$$\lim_{T\to 0} T^{2H}\log\mathbb{E}((\mathrm{VIX}_T-(a_1+\frac{1}{\Delta}\int_0^{\Delta}\xi_T(u)du)^{\frac{1}{2}}\,e^{xT^{\frac{1}{2}-H}})_+) \;=\; \lim_{T\to 0}T^{2H}\log\mathbb{E}((\mathrm{VIX}_T-\mathrm{VIX}_0\,e^{xT^{\frac{1}{2}-H}})_+)$$
$$= -J(2(a_1+\frac{1}{\Delta}\int_0^{\Delta}\xi_T(u)du)x)$$
$$= -J(2\mathrm{VIX}_0^2x)$$

where $J(.)$ is the same as in the main Theorem 2.1 in [FGS21b].

Proof. See Appendix A. ■

Corollary 2.2

$$\hat{\sigma}_{\mathrm{VIX}}(x) \;:=\; \lim_{T\to 0}\hat{\sigma}_{\mathrm{VIX}}(\mathrm{VIX}_0e^{xT^{H-\frac{1}{2}}},T) \;=\; \frac{|x|}{\sqrt{2J(2\mathrm{VIX}_0^2x)}}\,. \tag{4}$$

Expanding the final expression in Corollary 2.1, we find that

$$\hat{\sigma}_{\mathrm{VIX}}(x) \;=\; \frac{\nu\sqrt{V_0}}{\frac{1}{\Delta}\int_0^{\Delta}\xi_0(u)du+a_1}\,\frac{\Delta^{\alpha-1}}{2\alpha\Gamma(\alpha)} \;+\; \frac{\nu}{2\sqrt{V_0}\Gamma(2+\alpha)}x \;+\; O(x^2)$$
$$= \frac{\nu\sqrt{V_0}}{\mathrm{VIX}_0^2}\frac{\Delta^{\alpha-1}}{2\alpha\Gamma(\alpha)} \;+\; \frac{\nu}{2\sqrt{V_0}\Gamma(2+\alpha)}x \;+\; O(x^2)\,. \tag{5}$$

Remark 2.1 Since $a_1\geq 0$, we see that if the leading order term is fixed, adding jumps can only increase the skew term relative to the leading order term.

2.2 CGMY jumps and asymptotics for the SPX skew as $T\to 0$ in the Edgeworth regime

From here on we consider the special case when L is a generalized tempered stable (i.e. CGMY-type) process with $L_0=0$ and Lévy triple $(b,\sigma^2,\nu(x)=\frac{C_+e^{-Mx}}{x^{1+Y}}1_{x>0}+\frac{C_-e^{-G|x|}}{|x|^{1+Y}}1_{x<0})$ with $Y\in(1,2)$ for all $t>0$ which implies that

$$\phi_t^L(u) \;:=\; \mathbb{E}(e^{-iu(L_t-L_0)}) \;=\; e^{t\psi(u)} \;=\; e^{-iubt-\frac{1}{2}\sigma^2u^2t+C_+\Gamma(-Y)t[(M+iu)^Y-M^Y]+C_-\Gamma(-Y)t[(G-iu)^Y-G^Y]} \tag{6}$$

and $\mathbb{E}(e^{-pL_t}) = e^{V(p)t}$ for $p\in(-M,G)$, where $V(p)=\psi(-iu)$. Note that b is fixed by the martingale condition $V(1)=0$.

[2] Note this is conceptually similar to a simple displaced-diffusion model $dS_t=(\beta S_t+1-\beta)dW_t$ for S with $\beta\in(0,1)$ where S can go negative, but here we do not have this negativity problem since $a_1\geq 0$. This also means that a VIX put option with $K\leq a_1$ is worthless.

Remark 2.2 Corollary 2.10 and Remark 2.7 in [FSV21] describe a method for simulating L_t by changing to a measure under which L is a pure α-stable process, which allows for long-time stepping using the classical Chambers,Mallows&Stuck[CMS76] formula for sampling stable random variables.

We can replace the Y process in the main Theorem 2.1 in [FSV21] with our rough Heston V process (since the only special feature of the Y process required in [FSV21] is bivariate weak convergence of $(\frac{1}{\sqrt{t}}\hat{X}_t, (Y_t - \mathbb{E}(Y_t))/t^H)$ as $t \to 0$ to a joint Gaussian under Eq 17 in [FSV21], which is exact there even for $t > 0$ (but this is not required) but is also established in Appendix B of this article for $(\frac{1}{\sqrt{t}}\hat{X}_t, (V_t - \mathbb{E}(V_t))/t^H)$, and as in section 2.1 in [FSV21] we can extend from the case of bounded volatility to the true driftless rough Heston model (see Appendix C for details) and our S process is a martingale and the history term ζ in our context here (see just below Eq 5 in [FSV21]) is given by $\zeta(t) = \mathbb{E}(V_t) - V_0 = \xi_0(t) - V_0$ in our setup here, and the contribution from this term is contained in the error term in (7) if $\xi_0(t) = O(t^H)$ as $t \to 0$. Thus Theorem 2.1 (and Lemma 2.7) in [FSV21] still apply for the model in (2), which we now recall:

Proposition 2.3 *Let $\hat{\sigma}(K,T)$ denote the implied volatility of a European call option with strike K and maturity T at time zero. For the model in (2), if $H \in (1 - \frac{1}{2}Y, \frac{1}{2})$, $G > 1$ and $M > 1$, we have the following asymptotic behaviour for implied volatility at time zero in the Edgeworth regime:*

$$\hat{\sigma}(e^{z\sqrt{T}}, T) \;=\; \sqrt{V_0 + \sigma^2} + \frac{1}{\phi(\frac{z}{\sqrt{V_0+\sigma^2}})}(C_+A_+(z) + C_-A_-(z))T^{1-\frac{1}{2}Y} + o(T^{1-\frac{1}{2}Y}) \quad (7)$$

for $z \in \mathbb{R}$ as $T \to 0$ and

$$A_\pm(z) \;=\; \frac{1}{\pi}\int_0^\infty \mathrm{Re}[e^{-iuz}e^{-\frac{1}{2}(V_0+\sigma^2)u^2}\Gamma(-Y)(\pm iu)^{Y-2}]du$$

where ϕ is the standard Normal density.

Remark 2.3 $\frac{1}{\phi(\frac{z}{\sqrt{V_0+\sigma^2}})}(C_+A_+(z) + C_-A_-(z))$ is non-linear in z and does not vanish at $z = 0$ in general (see e.g. Figure 6 in [FSV21] for a plot), and C^+ and M relate to negative jumps for X and C^- and G relate to positive jumps, since we are subtracting L in (2). Note that the log stock price X_t does not satisfy a small-time LDP as $t \to 0$ (even if re-scaled).

Remark 2.4 Note that a_1 depends on M and G as well, which do not show up in the leading order Edgeworth asymptotics in (7).

Hence the addition of the CGMY-type jump process L allows the asymptotic SPX and VIX short time smiles to have power-law skews with different H-values since the effective H as $T \to 0$ for the former is now $1 - Y/2$ (due to jumps alone, see (7)) and the latter is $H = \alpha - \frac{1}{2}$ (due to V alone), so (aside from (7) depending on V_0), we have effectively de-coupled the short-time behaviour of the SPX and VIX skews.

2.3 Exact calibration to the at-the-money skew for SPX and VIX as $T \to 0$

We let $\hat{\sigma}^{\mathrm{mkt}}_{\mathrm{VIX}}(K,T)$ denote the *market* implied volatility of a VIX call option with strike K and maturity T. If there exists an $\alpha = H + \frac{1}{2}$ such that $\lim_{T\to 0}\hat{\sigma}^{\mathrm{mkt}}_{\mathrm{VIX}}(\mathrm{VIX}_0 e^{xT^{H-\frac{1}{2}}}, T)$ has a non-trivial limit which is differentiable in x, then we can choose V_0 and ν so that the market observed behaviour of VIX options as $T \to 0$ is consistent with (5). We can then choose parameters so that (7) is consistent with the observed behaviour of SPX options in the market. Specifically choose σ to fit the observed behaviour of the limiting at-the-money implied vol $\lim_{T\to 0}\hat{\sigma}^{\mathrm{mkt}}(1,T)$, and then solve for C^+, C^- and Y to match the observed behaviour of the at-the-money correction for European options: $\lim_{T\to 0}\frac{\hat{\sigma}^{\mathrm{mkt}}(1,T) - \hat{\sigma}^{\mathrm{mkt}}(1,0)}{T^{1-Y/2}}$ and the skew term: $\lim_{z\to 0}\lim_{T\to 0}\frac{\hat{\sigma}^{\mathrm{mkt}}(e^{z\sqrt{T}},T) - \hat{\sigma}^{\mathrm{mkt}}(1,T)}{zT^{1-Y/2}}$, assuming we can find a $Y \in (1,2)$ such that both these limits are finite (and at least one is non-zero). This leads to the following two equations for $C_\pm$:

$$\sqrt{2\pi}(C^+A_+(0) + C^-A_-(0)) \;=\; \lim_{T\to 0}\frac{\hat{\sigma}^{\mathrm{mkt}}(1,T) - \hat{\sigma}^{\mathrm{mkt}}(1,0)}{T^{1-Y/2}} \quad (8)$$

$$\sqrt{2\pi}(C^+A'_+(0) + C^-A'_-(0)) \;=\; \lim_{z\to 0}\lim_{T\to 0}\frac{\hat{\sigma}^{\mathrm{mkt}}(e^{z\sqrt{T}},T) - \hat{\sigma}^{\mathrm{mkt}}(1,T)}{zT^{1-Y/2}} \quad (9)$$

obtained by computing $\frac{1}{\phi(\frac{z}{\sqrt{V_0+\sigma^2}})}(C_+A_+(z)+C_-A_-(z))$ and its derivative at $z=0$. The last two equations are linear in C_+ and C^- and hence can be solved explicitly if the Wronskian of $A_+(z)$ and $A_-(z)$ is non-zero at $z=0$, and we can verify that this is indeed the case since

$$\begin{aligned} A_\pm(z) &= \frac{1}{\pi}\int_0^\infty \mathrm{Re}[e^{-iuz}e^{-\frac{1}{2}(V_0+\sigma^2)u^2}\Gamma(-Y)(\pm iu)^{Y-2}]du \\ &= \frac{1}{\pi}\int_0^\infty \mathrm{Re}[e^{-iuz}e^{-\frac{1}{2}(V_0+\sigma^2)u^2}\Gamma(-Y)e^{\pm i\pi(Y-2)}u^{Y-2}]du \\ &= \frac{1}{\pi}\int_0^\infty \cos(-uz\pm\pi(Y-2))e^{-\frac{1}{2}(V_0+\sigma^2)u^2}\Gamma(-Y)u^{Y-2}du \end{aligned}$$

which implies that

$$\begin{aligned} A'_\pm(z) &= \frac{1}{\pi}\int_0^\infty \mathrm{Re}[u\sin(-uz\pm\pi(Y-2))e^{-\frac{1}{2}(V_0+\sigma^2)u^2}\Gamma(-Y)u^{Y-2}]du \\ A_\pm(0) &= \frac{1}{\pi}\int_0^\infty \mathrm{Re}[\cos(\pi(Y-2))e^{-\frac{1}{2}(V_0+\sigma^2)u^2}\Gamma(-Y)u^{Y-2}]du \\ A'_\pm(0) &= \pm\frac{1}{\pi}\int_0^\infty \mathrm{Re}[u\sin(\pi(Y-2))e^{-\frac{1}{2}(V_0+\sigma^2)u^2}\Gamma(-Y)u^{Y-2}]du \end{aligned}$$

i.e. $A_+(0)=A_-(0)$ and $A'_+(0)=-A'_-(0)$.

We also have to make the model consistent with the observed variance term structure (i.e. prices of variance swaps at all maturities implied by European option prices) up to some maturity $T^*\in(0,\Delta)$, for which

$$\begin{aligned} -2\mathbb{E}(\log\frac{S_t}{S_0}) &= \int_0^t \xi_0(u)du\,+\,a_1t \\ \Rightarrow\quad -\frac{\partial}{\partial t}2\mathbb{E}(\log\frac{S_t}{S_0}) &= \xi_0(t)\,+\,a_1 \\ \Rightarrow\quad -\frac{\partial}{\partial t}2\mathbb{E}(\log\frac{S_t}{S_0})_{t=0} &= \xi_0(0)\,+\,a_1 \;=\; V_0\,+\,a_1\,. \end{aligned} \tag{10}$$

a_1 depends on G and M, which must be chosen so the calibrated value $\xi_0(0)=V_0$ here matches our calibrated value of V_0 from above; then (10) uniquely determines $\xi_0(t)$ for $t>0$ as well (and this includes being consistent with the market value of VIX_0 which corresponds to $t=\Delta$), which we have implicitly used above for the VIX smile calibration).

We now discuss the range of admissible values for a_1. Imposing the martingale condition $V(1)=0$ we find that

$$\begin{aligned} a_1 \;=\; a_1(G,M) &= \sigma^2\,-\,2C_+(M^Y-(1+M)^Y+M^{Y-1}Y)\Gamma(-Y) \\ &+\;2C_-((G-1)^Y-G^Y-G^{Y-1}Y)\Gamma(-Y)\,. \end{aligned}$$

Then $a_1(1,1)>0$ and $a_1(G,M)\to\sigma^2$ as $M,G\to\infty$, and we have the following monotonicity result:

Lemma 2.4 *$a_1(G,M)$ is continuous and strictly decreasing in M and G.*

Proof.

$$\frac{\partial a_1}{\partial M} = -2C_+(YM^{Y-1}-Y(1+M)^{Y-1}+Y(Y-1)M^{Y-2})\Gamma(-Y) \tag{11}$$

Now consider the function $g_M(K)=(K+M)^{Y-1}$. This function is concave so

$$(K+M)^{Y-1} \;\le\; M^{Y-1}\,+\,(Y-1)M^{Y-2}K$$

Setting $K=1$ we see that

$$(1+M)^{Y-1} \;\le\; M^{Y-1}\,+\,(Y-1)M^{Y-2}$$

and the result follows by comparing this to (11). We use a similar argument for the term involving G. ■

Since Proposition 2.3 requires that $G,M>1$, we see that there is an interval (σ^2,a_1^*) of admissible a_1-values, where $a_1^*=a_1(1,1)=\sigma^2+2(C_+(2^Y-Y-1)+C_-(Y-1))\Gamma(-Y)$ for Y in the admissible range $Y\in(1,2)$. If the calibrated V_0-value is such that $-\frac{\partial}{\partial t}2\mathbb{E}(\log\frac{S_t}{S_0})_{t=0}-V_0\notin(\sigma^2,a_1^*]$

(see (10)) then it means the model is mis-specified, and this ultimately puts an upper and lower bound on the admissible ratio of the first to the zeroth order term in (5).

Of course in practice we cannot perfectly extract all the limiting quantities on the right hand side of (8) and (9) from the market since only a finite number of options are traded, and there are real-life issues with noisy data, bid-offer spreads, day count/weekend conventions etc., but one can approximate the quantities on the right with finite differences to incorporate information from options with small but non-zero maturity. These results can also be used to make smart initial guesses for conventional and deep learning-based calibration schemes.

J.Guyon has remarked[3] has remarked that fitting large SPX short-end skew and (comparatively) small VIX short-end implied volatility is the main challenge in joint fitting of SPX and VIX smiles, so this gives a concrete result in this direction.

Remark 2.5 This method (and model) may avoid potential problems of overfitting using e.g. Schrödinger bridge-type model (see [HL19] and Guyon[Guy20b]) which is a time-inhomogenous diffusion model which is Markovian in S_t and V_t. Our model is entirely plausible as a natural time-homogenous extension of the rough Heston model

3 Short-maturity skew and other properties of the quadratic rough Heston model

In this section we discuss some peculiarities of the quadratic rough Heston model introduced in [GJR20] for which they report some encouraging results for joint calibration to SPX and VIX smiles for small and medium maturity options. For a log stock price process X_t, the model is defined as

$$\begin{cases} dX_t &= -\frac{1}{2}V_t dt + \sqrt{V_t}dW_t\,, \\ Z_t &= \frac{1}{\Gamma(\alpha)}\int_0^t (t-s)^{\alpha-1}(\lambda(\theta(s)-Z_s)ds + \nu\sqrt{V_s}dW_s) \end{cases} \tag{12}$$

for $H=\alpha-\frac{1}{2}\in(0,\frac{1}{2})$, where $V_t=a(Z_t-b)^2+c$, $\nu,\lambda,a,b,c>0$ and W is a one-dimensional Brownian motion. We note that the model has the (potentially slightly unrealistic) feature that $V_t\geq c>0$, so

$$\text{VIX}_T \;=\; (\frac{1}{T}\int_T^{T+\Delta}\mathbb{E}_T(V_u)du)^{\frac{1}{2}} \;\geq\; \sqrt{c}$$

which implies that the VIX implied volatility is zero for strikes $K\leq\sqrt{c}$ at all strikes, since in this case $\mathbb{E}((\text{VIX}_T-K)^+)=\mathbb{E}(\text{VIX}_T)-K$, i.e. there is no time-value to the option[4], and the same is true for our Rough Heston-CGMY model in (2), since the a_1 constant in (3) is positive.

Formally expanding V_t around V_0, we see that $V_t\approx const.+\beta(Z_t-Z_0)=\alpha+\beta Z_t$ where $\beta=2a(Z_0-b)$ and $\alpha=ab^2+c-aZ_0^2$ (note with mild abuse of notation we are not defining α to be $H+\frac{1}{2}$ in this subsection), so we expect the model to be locally approximated over small time intervals by the following affine model:

$$\begin{cases} dX_t &= -\frac{1}{2}(\alpha+\beta Z_t)dt + \sqrt{\alpha+\beta Z_t}dW_t\,, \\ Z_t &= Z_0 + \int_0^t (t-s)^{H-\frac{1}{2}}\frac{\nu}{\Gamma(H+\frac{1}{2})}\sqrt{\alpha+\beta Z_s}dW_s \end{cases}$$

and setting $Y_t:=\beta Z_t$, we see that

$$\begin{cases} dX_t &= -\frac{1}{2}(\alpha+Y_t)dt + \sqrt{\alpha+Y_t}dW_t\,, \\ Y_t &= Y_0 + \int_0^t (t-s)^{H-\frac{1}{2}}\frac{\beta\nu}{\Gamma(H+\frac{1}{2})}\sqrt{\alpha+Y_s}dW_s \end{cases}$$

and further defining $\tilde{Y}_t:=\alpha+Y_t$ we see that

$$\begin{cases} dX_t &= -\frac{1}{2}\tilde{Y}_t dt + \sqrt{\tilde{Y}_t}dW_t\,, \\ \tilde{Y}_t &= \tilde{Y}_0 + \int_0^t (t-s)^{H-\frac{1}{2}}\frac{\beta\nu}{\Gamma(H+\frac{1}{2})}\sqrt{\tilde{Y}_s}dW_s \end{cases}$$

which is now just a standard rough Heston model with ν replaced by $\beta\nu$, and note that we have ignored the drift terms of V since they not affect the small-time asymptotics at the order we consider here.

[3]in private communication

[4]We thank Alan Lewis for pointing out this behaviour for Markovian models

Then formally appealing to Corollary 3.4 in [FSV21] (see also page 16 there for the definition of the b parameter) for the model in (12), we have the following asymptotic behaviour for the implied volatility in the small-maturity limit:

$$\frac{1}{T^H}(\sigma_{\text{impl}}(z\sqrt{T},T)-\sigma_{\text{impl}}(0\sqrt{T},T)) \quad = \quad \frac{b_1\beta\nu}{V_0^{\frac{1}{2}}}z\,+\,o(T) \quad = \quad \frac{2a(Z_0-b)\nu}{2\Gamma(\frac{5}{2}+H)V_0^{\frac{1}{2}}}z\,+\,o(T) \qquad (13)$$

as $T\to 0$ for $z\in\mathbb{R}$ where $b_1=\frac{1}{2\Gamma(2+H+\frac{1}{2})}=\frac{1}{2\Gamma(\frac{5}{2}+H)}$, so we see that the skew term is positive if $Z_0>b$, and we have numerically corroborated (13) using Monte Carlo methods (see also Figure 2 below). We expect to get the same result even if we are conditioning on a previous history of V, since from Eq 2 in [GJR20] this history is an a.s. bounded drift term which will not affect the asymptotic behaviour at the order we are interested in (13), which means even if the model is calibrated at time zero with negative skew, the short-maturity skew can flip to positive at some future time.

To sample VIX_T under the quadratic rough Heston model, one uses that $\xi_t(u)=\mathbb{E}_t(V_u)$ satisfies a linear VIE with (random) history terms which involve stochastic integrals whose values are known at time t which can be solved explicitly in term of the resolvent of the second kind of the kernel (see Theorem 6.1 in [Rom22] for details).

[GJR20] and [RZ21] report calibrated H-values of .01, but Monte Carlo is notoriously innacurate for such low H-values particularly for extreme values of the correlation and large absolute log-moneyness values (in e.g. the $T=.033$ SPX smile plot in [GJR20]), see e.g.[Gath21] and [Rom22] for more on these issues), and it is unknown whether S is a true martingale (ρ is essentially ± 1 for the quadratic rough Heston model depending on the current value of Z_t, and it is known that for many Markovian models S is not a true martingale for positive ρ, see e.g. [AP07],[LM07],[Jour04].

The finite-dimensional Markov approximation used in [RZ21] for the quadratic rough Heston model (see also [BB23] and [MW21]) uses n exponentials to approximate the fractional kernel (see Appendix A in [RZ21] to see how the exponentials are chosen), and one then has to solve an $n\times n$ system of linear inhomogenous ODEs with random initial condition to sample VIX_T^2 under the approximating model which is not mentioned in the article (see also [Rom22]), but even just for the SPX smile, the Markov approximation also typically exhibits huge sample variance for implied volatility using Monte Carlo when $H=.01$ even for 10million sample paths and 500-800 time steps. The approximating model is not rough, so in particular will not exhibit power-law skew in the $T\to 0$ limit, and obtaining power-law skew is one of the principal reasons for using rough volatility models).

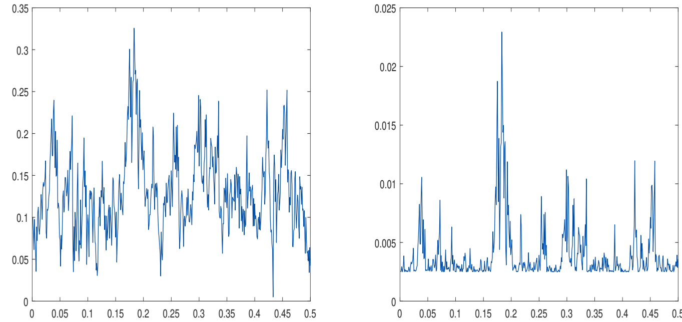

Figure 1: On the left here we see a Monte Carlo sample path of Z using a Volterra Euler-type scheme and the corresponding sample path for V (second plot) for $t \in [0, 0.5]$ for parameters $\alpha = .51$, so $H = \alpha - .5 = .01$, $\eta = 1$, $\lambda = 1.2$, $Z_0 = .1$, $\theta = 0.0835$; $a = .384$, $b = .095$, $c = .0025$.

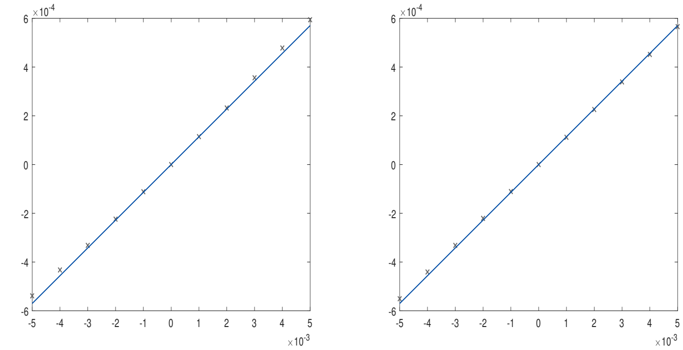

Figure 2: Here we have plotted the Monte Carlo values for $\frac{\sigma_{\mathrm{impl}}(z\sqrt{T},T)-\sigma_{\mathrm{impl}}(0,T)}{T^H}$ (grey crosses) verses the theoretical value for the skew: $\frac{b_1 \beta \nu}{V_0^{\frac{1}{2}}} z$, as a function of z for $T = .00001$ (left plot) and $T = .001$ (right plot) and $\alpha = .75$, $Z_0 = .12$, $\theta = .1$ and a, b, c the same as the previous plot with 400 time steps and 250,000 sample paths

References

[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*.

[AP07] Andersen, L.B.G., Piterbarg, V.V., "Moment Explosions in Stochastic Volatility Models", *Finance and Stochastics*, 11 (1), 29-50, 2007.

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

[BB23] Bayer, C. and S.Breneis, "Markovian approximations of stochastic Volterra equations with fractional kernel", *Quantitative Finance*, Volume 23, 2023 - Issue 1

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

[BL00] Boyarchenko, S, and S.Levendorski, "Option pricing for truncated Lévy processes", *International Journal of Theoretical and Applied Finance*, 3(3), 2000, pp. 549–552

[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.

[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.

[FGS21b] Forde, M., S.Gerhold and B.Smith, "Small-time VIX smile for the Rough Heston model", with S.Gerhold and B.Smith (2021), preprint.

[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.

[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).

[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.

[GLOW21] Guo, I., G.Loeper, J.Obloj, S.Wang, "Optimal transport for model calibration", preprint

[GLOW19] Guo, I., G.Loeper, S.Wang, "Calibration of Local-Stochastic Volatility Models by Optimal Transport", preprint

[Gath21] Gatheral, J., "Efficient simulation of affine forward variance models", preprint

[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

[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

[JP20] Jacquier, A. and A.Pannier, “Large and Moderate Deviations for Stochastic Volterra Systems”, preprint

[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.

[Jour04] Jourdain, B., “Loss of Martingality in Asset Price Models with Lognormal Stochastic Volatility”, preprint

[LM07] Lions, P.L., Musiela, M.: “Correlations and bounds for stochastic volatility models”, *Annales de l’Institut Henri Poincaré*, 24 (1), 1-16, 2007

[MW21] Jingtang Ma, and Haofei Wu, “A fast algorithm for simulation of rough volatility models”, *Quantitative Finance*, DOI: 10.1080/14697688.2021.1970213

[RZ21] Mathieu Rosenbaum, Jianfei Zhang “Deep calibration of the quadratic rough Heston model”, preprint, 2021

[Rom22] Rømer, S. “Hybrid multifactor scheme for stochastic Volterra equations with completely monotone kernels”, preprint

[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 Asymptotics for VIX call options

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

$$(\mathrm{VIX}_T^2)^q = (a_1 + \frac{1}{\Delta}\int_T^{T+\Delta} \xi_T(u)du)^q = (\frac{1}{\Delta}\int_T^{T+\Delta} (a_1 + \xi_T(u)du))^q \leq \frac{1}{\Delta}\int_T^{T+\Delta} (a_1 + \xi_T(u))^q du$$

and hence

$$\begin{aligned}\mathbb{E}(\mathrm{VIX}_T^{2q}) &\leq \frac{1}{\Delta}\int_T^{T+\Delta} \mathbb{E}((a_1 + \xi_T(u))^q)du = \frac{1}{\Delta}\int_T^{T+\Delta} \mathbb{E}(\,\mathbb{E}(a_1 + V_u|\mathcal{F}_T)^q)du \\ &\leq \frac{1}{\Delta}\int_T^{T+\Delta} \mathbb{E}((a_1 + V_u)^q)du \qquad \text{(D-1)}\end{aligned}$$

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 $\mathrm{VIX}_0^2 = a_1 + \frac{1}{\Delta}\int_T^{T+\Delta}\xi_T(u)du$ 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

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

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

$$
\begin{aligned}
&T^{2H}\log\mathbb{E}((\mathrm{VIX}_T - \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H}))_+) \\
\leq\; &\frac{T^{2H}}{q}\log\mathbb{E}[(\mathrm{VIX}_T - \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H}))_+^q] + T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T \geq \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H})) \\
\leq\; &\frac{T^{2H}}{q}\log\mathbb{E}(\mathrm{VIX}_T^q) + T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T \geq \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H})) \\
\leq\; &\frac{T^{2H}}{q}\log(\mathbb{E}(\mathrm{VIX}_T^{2q})^{\frac{1}{2}}) + T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T \geq \mathrm{VIX}_0\,(1+xT^{\frac{1}{2}-H})) \\
\leq\; &\frac{T^{2H}}{q}\frac{1}{2}\log(\frac{1}{\Delta}\int_T^{T+\Delta}\mathbb{E}((a_1+V_u)^q)du) + T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T^2 \geq \mathrm{VIX}_0^2(1+xT^{\frac{1}{2}-H})^2) \\
&\text{(by (D-1))} \\
\leq\; &\frac{T^{2H}}{q}\frac{1}{2}\log(\frac{1}{\Delta}\int_T^{T+\Delta}(a_1+\mathbb{E}(V_u^q)^{\frac{1}{q}})^q du + T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T^2 \geq \mathrm{VIX}_0^2(1+2xT^{\frac{1}{2}-H})) \\
&\text{(using Minkowski applied to } \mathbb{E}((a_1+V_u)^q)) \\
\leq\; &\frac{T^{2H}}{q}\frac{1}{2}\log(a_1+c_{q,T}^{\frac{1}{q}})^q + T^{2H}(1-\frac{1}{q})\log\mathbb{P}(\mathrm{VIX}_T^2 \geq \mathrm{VIX}_0^2(1+2xT^{\frac{1}{2}-H})
\end{aligned}
$$

for some finite constant $c_{q,T}$ depending on q and T, where we have used Lemma C.1 in [FGS21] 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

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

B Bivariate weak convergence

Proposition B.1 *Let $\hat{X}_t := \sqrt{V_0}(\rho W_t + \bar{\rho}B_t)$. Then $(X_t/\sqrt{t}, (V_t-\xi_0(t))/t^H)$ and $(\hat{X}_t/\sqrt{t}, (V_t-\xi_0(t))/t^H)$ both tend weakly to the same bivariate Gaussian as $t \to 0$.*

Proof.

$$\mathbb{E}(e^{pX_t+u(V_t-\xi_0(t))}) \quad = \quad e^{\int_0^t \xi_0(t-s)(\frac{1}{2}p^2 - \frac{1}{2}p + p\rho\nu\psi(p,u,s)+\frac{1}{2}\nu^2\psi(p,u,s)^2)ds}$$

for $t \in [0, T^*_\psi(p,u))$, where $\psi(p,u,t)$ satisfies

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

and $T^*_\psi(p,u) > 0$ is the explosion time for ψ. Then

$$\mathbb{E}(e^{\frac{p}{\sqrt{\varepsilon}}X_{\varepsilon t}+\frac{u}{\varepsilon^H}(V_{\varepsilon t}-\xi_0(t))}) \quad = \quad e^{\int_0^{\varepsilon t} \xi_0(\varepsilon t-s)(\frac{1}{2}\frac{p^2}{\varepsilon} - \frac{1}{2}\frac{p}{\sqrt{\varepsilon}} + \frac{p}{\sqrt{\varepsilon}}\rho\nu\psi(\frac{p}{\sqrt{\varepsilon}},\frac{u}{\varepsilon^H},s)+\frac{1}{2}\nu^2\psi(\frac{p}{\sqrt{\varepsilon}},\frac{u}{\varepsilon^H},s)^2)ds}$$

and

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

for $t \in [0, \frac{1}{\varepsilon}T^*_\psi(\frac{p}{\sqrt{\varepsilon}}, \frac{u}{\varepsilon^H}))$. Then $\psi^\varepsilon(p,u,t) = \sqrt{\varepsilon}\psi(p,u,\varepsilon t)$ satisfies

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

for $t \in [0, \frac{1}{\varepsilon}T^*_\psi(\frac{p}{\sqrt{\varepsilon}}, \frac{u}{\varepsilon^H}))$, and

$$\begin{aligned}
\mathbb{E}(e^{\frac{p}{\sqrt{\varepsilon}}X_{\varepsilon t}+\frac{u}{\varepsilon^H}(V_{\varepsilon t}-\xi_0(t))}) \quad &= \quad e^{\int_0^{\varepsilon t} \xi_0(\varepsilon t-s)(\frac{1}{2}\frac{p^2}{\varepsilon} - \frac{1}{2}\frac{p}{\sqrt{\varepsilon}} + \frac{p}{\sqrt{\varepsilon}}\rho\nu\psi(p,u,s)+\frac{1}{2}\nu^2\psi(p,u,s)^2)ds} \\
&= \quad e^{\varepsilon\int_0^t \xi_0(\varepsilon t-\varepsilon s)(\frac{1}{2}\frac{p^2}{\varepsilon} - \frac{1}{2}\frac{p}{\sqrt{\varepsilon}} + \frac{p}{\sqrt{\varepsilon}}\rho\nu\psi(p,u,\varepsilon s)+\frac{1}{2}\nu^2\psi(p,u,\varepsilon s)^2)ds} \\
&= \quad e^{\int_0^t \xi_0(\varepsilon t-\varepsilon s)(\frac{1}{2}p^2 - \frac{1}{2}p\sqrt{\varepsilon}+p\rho\nu\psi^\varepsilon(p,u,s)+\frac{1}{2}\nu^2\psi^\varepsilon(p,u,\varepsilon s)^2)ds}.
\end{aligned}$$

From Theorem 13.1.1 i) in [GLS90] (see the paragraph above Eq 12 in [FGS21] for details on the use of part i) of Theorem 13.1.1 in [GLS90]), we know that $\psi^\varepsilon(p,u,t)$ tends uniformly to $uc_\alpha t^{\alpha-1}$ as $\varepsilon \to 0$ on any compact interval, and since $\xi_0(.)$ is continuous (and hence bounded on any compact interval) we see that

$$\mathbb{E}(e^{\frac{p}{\sqrt{\varepsilon}}\tilde{X}^\varepsilon_t+\frac{u}{\varepsilon^H}(V^\varepsilon_t-\xi_0(t))}) \quad \to \quad e^{V_0(\frac{1}{2}p^2 t + \rho\nu\rho u\frac{t^\alpha}{\Gamma(\alpha+1)} + \frac{1}{2}\nu^2 u^2\frac{t^{2\alpha-1}}{\Gamma(\alpha)^2(2\alpha-1)})}$$

by the bounded convergence theorem. 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.

Similarly

$$\mathbb{E}(e^{\frac{p}{\sqrt{\varepsilon}}\hat{X}^\varepsilon_t+\frac{u}{\varepsilon^H}(V^\varepsilon_t-\xi_0(t))}) \quad = \quad e^{\int_0^t \xi_0(T-s)(\frac{1}{2}p^2 - \frac{1}{2}p\sqrt{\varepsilon}+\varepsilon^H p\rho\nu\psi^\varepsilon(p,u,s)+\frac{1}{2}\varepsilon^{2H}\nu^2\psi^\varepsilon(p,u,s)^2)ds}$$

where $\psi^\varepsilon(p,u,t)$ satisfies

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

with $\psi^\varepsilon(p,0) = 0$, and the rest of the proof follows using the same arguments as above. ■

C Extending to unbounded volatility

This appendix is a minor variation of Section 2.1 in [FSV21] adapted for our purposes here. We first let $\underline{V}_t := \min_{0\le s\le t}V_s$ and $\bar{V}_t := \max_{0\le s\le t}V_s$, and we consider the stochastic Volterra system

$$\begin{aligned}
d\hat{X}_t \quad &= \quad -\frac{1}{2}\sigma(\hat{V}_t)^2 dt + \sigma(\hat{V}_t)(\rho dW_t + \bar{\rho}dB_t) \\
\hat{V}_t \quad &= \quad \xi_0(t) + \int_0^t (t-s)^{H-\frac{1}{2}}\sigma(\hat{V}_s)dW_s
\end{aligned}$$

where $\sigma \in C_b^2$ as in [FSV21] and bounded away from zero with $\sigma(y)=\nu\sqrt{y}$ for $y\in[a,b]$ for some $a\in(0,V_0)$ and $b>V_0$, so in particular σ is bounded and Lipshitz. Then $V_t=\hat{V}_t$ for $t\in[0,\tau_a\wedge\tau_b)$ where $\tau_a=\inf\{t: V_t=a\}$ (and similarly for τ_b) for any $a\in(0,V_0)$ since we have pathwise uniqueness for $\hat{V}$ (see section 4.2.3 in [JP20] and discussion in section 2 in [FGS21]) and from Proposition 4.3 in [JP20] (or Theorem 3.11 in [Zha08]), we know that $\hat{V}_{t(.)}$ satisfies the large deviation principle on $C_0[0,1]$ as $t\to 0$ with speed $1/t^{2H}$ and rate function

$$I(\phi) \;=\; \frac{1}{2}\int_0^1 \Big(\frac{D^\alpha(\phi(.)-\phi(0))(t)}{\sigma(\phi(t))}\Big)^2 dt$$

if $\phi\in I_{V_0}^{H+\frac{1}{2}}(L^1)$, otherwise $I(\phi)=\infty$. Now let $A:=\{\underline{V}_t>a\}\cap\{\bar{V}_t<b\}$ and set $\hat{S}_t=e^{\hat{X}_t}$, where

$$d\hat{X}_t \;=\; -\frac{1}{2}\hat{V}_t dt + \sqrt{\hat{V}_t}(\rho dW_t+\bar{\rho}dW_t^\perp) - dL_t$$

and $\hat{X}_0=0$ i.e. the same stock price process (defined on the same probability space) but with V_t replaced with $\hat{V}_t$. Then for all $\varepsilon>0$ we have

$$\begin{aligned}
\mathbb{E}((e^{z\sqrt{t}}-S_t)^+) &= \mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+1_A) + \mathbb{E}((e^{z\sqrt{t}}-S_t)^+1_{A^c})\\
&\le \mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+) + e^{z\sqrt{t}}\mathbb{P}(A^c)\\
&\le \mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+) + e^{z\sqrt{t}}e^{-\frac{1}{t^{2H}}(-\varepsilon+J(a,b))}
\end{aligned}$$

for t sufficiently small, where $J(a,b)=\inf_{\phi\in C_{V_0}([0,1]):\underline{\phi}(1)\le a \text{ or } \bar{\phi}(1)\ge b} I(\phi)$, and we have used the upper bound implied by the aforementioned LDP in the final line. For this to be useful, we need to check that $J(a,b)>0$. To this end, let $\mathcal{I}: H_0^1\to C_0([0,1])$ be the Itô map which takes h to the solution to $\phi(t)=V_0+\int_0^t\frac{1}{\Gamma(\alpha)}(t-s)^{H-\frac{1}{2}}\dot{h}_s\sigma(\phi(s))ds$ (recall this solution is unique since σ is bounded and Lipschitz). Then

$$\begin{aligned}
|\mathcal{I}(h)(t)-V_0| \;=\; |\phi(t)-V_0| \;=\; |\int_0^t c_\alpha(t-s)^{H-\frac{1}{2}}\sigma(h(s))\dot{h}_s ds| &\le |\int_0^t c_\alpha(t-s)^{H-\frac{1}{2}}\sigma(b)\dot{h}_s ds|\\
&\le const.\times\frac{1}{2}\Big(\int_0^t \dot{h}_s^2 ds\Big)^{\frac{1}{2}}\\
&= const.\times J(\phi)
\end{aligned}$$

where we have used Cauchy-Schwarz in the final line. If $J(a,b)=0$ then for all $\varepsilon>0$ there exists a $\phi_\varepsilon\in C_{V_0}([0,1]$ with $\underline{\phi}_\varepsilon(1)\le a$ or $\bar{\phi}_\varepsilon(1)\ge b$ and $I(\phi_\varepsilon)\le\varepsilon$ for all $\varepsilon>0$, which (from the bound immediately above) implies that the sup norm of $\phi_\varepsilon-V_0$ is less than $(b-V_0)\wedge(V_0-a)$ for ε sufficiently small), which is a contradiction. Thus $J(a,b)>0$.

Similarly

$$\begin{aligned}
\mathbb{E}((e^{z\sqrt{t}}-S_t)^+) &\ge \mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+1_A)\\
&= \mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+) - \mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+1_{A^c})\\
&\ge \mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+) - e^{z\sqrt{t}}e^{-\frac{1}{t^{2H}}(J(a,b)-\varepsilon)}.
\end{aligned}$$

Thus

$$\begin{aligned}
\frac{1}{\sqrt{t}}\mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+) - \frac{e^{z\sqrt{t}}}{\sqrt{t}}e^{-\frac{1}{t^{2H}}(J(a,b)-\varepsilon)} &\le \frac{1}{\sqrt{t}}\mathbb{E}((e^{z\sqrt{t}}-S_t)^+)\\
&\le \frac{1}{\sqrt{t}}\mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+) + \frac{e^{z\sqrt{t}}}{\sqrt{t}}e^{-\frac{1}{t^{2H}}(J(a,b)-\varepsilon)}.
\end{aligned}$$

$\hat{S}$ satisfies the conditions of the main Theorem 2.1 in [FSV21], so we have a small-t expansion for $\frac{1}{\sqrt{t}}\mathbb{E}((e^{z\sqrt{t}}-\hat{S}_t)^+)$, and for t small, $\frac{1}{\sqrt{t}}e^{z\sqrt{t}}e^{-\frac{1}{t^{2H}}(\inf_{y\ge a}J(y)-\varepsilon)}$ is higher order than the error term in the main Theorem 2.1 in [FSV21], so the same expansion holds for $\frac{1}{\sqrt{t}}\mathbb{E}((e^{z\sqrt{t}}-S_t)^+)$.