rHeston

Nodari Alessandro & Proserpio Lorenzo

As we have seen the skew of the at the money options in the Heston model does not represent well the skew of real-world at the money options. This, among other reasons, made the researcher and practitioner to think about more sophisticated models. As observed by Gatheral, Jaisson and Rosenbaum the log-volatility in the market behaves like a fractional Brownian motion with small Hurst parameter, so the volatility is rough.

1 Stylized empirical facts

Heston model reproduces several important features of low frequency price data, provides quite reasonable dynamics for the volatility surface and it can be calibrated efficiently. If we want to surpass that we have to build a model which can reproduce the stylized facts of modern electronic markets in the context of high frequency trading. In practice each market behaves tick-by-tick, indeed we receive an update in price by market-makers whenever there is a trade and the movement is discrete and at least of one tick (usually 1 cent). There are 4 main stylized facts that we can observe in market data:

1. Markets are highly endogenous, as showed by Bouchad. This means that most of the orders have no real economic motivation, but are simply the reaction of algorithms to other orders.

2. Markets at high frequency are much more efficient than at lower frequencies, this means that it is much more difficult to find profitable statistical arbitrage strategies.

3. There is some asymmetry in the liquidity on the bid and the ask side of the order book. Indeed, a market-maker is likely to raise the price by less following a buy order than to lower the price following the same size sell order, as seen by Brunnermeier and Pedersen. This is mostly due to the fact that hedging the first position is easier than the second and that market-makers have usually some inventory.

4. A large proportion of transaction is due to big orders, called metaorders, which are not executed at once, but split in time. Indeed, one of the most challenging part of every trading strategies is to execute it in large volumes without moving changing to much the state of the market.

2 Building the model

As in El Euch, Fukasawa & Rosenbaum we will start building a model from Hawkes processes, then slowly including the stylized fact mentioned in the last paragraph and showing that the long-term dynamic of this model will lead to a rough Heston model at the macroscopic scale, in which the leverage effect is still represented.

2.1 Hawkes Processes

Hawkes processes are point processes which are said to be self-exciting, in the sense that the instantaneous jump-probability depends on the location of the past events. In particular we will focus on a bivariate Hawkes process, $(N_t^+, N_t^-)_{t\geq 0}$, where N_t^+ is the number of upward jumps of one tick and N_t^- is the number of downward jumps of one tick, both in the interval $[0, t]$. The probability to get one-tick upward jump in a time $\mathrm{d}t$ is given by $\lambda_t^+ \mathrm{d}t$, viceversa by λ_t^-. The array $(\lambda_t^+, \lambda_t^-)$ is called intensity of the process and it is of the form:

$$\begin{pmatrix}\lambda_t^+ \\ \lambda_t^-\end{pmatrix} = \begin{pmatrix}\mu^+ \\ \mu^-\end{pmatrix} + \int_0^t \begin{pmatrix}\phi_1(t-s) & \phi_3(t-s) \\ \phi_2(t-s) & \phi_4(t-s)\end{pmatrix} \cdot \begin{pmatrix}\mathrm{d}N_s^+ \\ \mathrm{d}N_s^-\end{pmatrix}$$

where μ^+ and μ^- are positive constants and the components of the matrix are positive and locally integrable functions. The process of the prices P_t is given by the difference between the number of upward jumps from time 0 and the number of downward jumps, so:

$$P_t = N_t^+ - N_t^-$$

Each component of the intensity can be decomposed into three parts, for example λ_t^- can be decomposed in:

- μ_t^- which corresponds to the probability that the price will go down because of exogenous reasons;
- $\int_0^t \phi_2(t-s)\mathrm{d}N_s^+$ which corresponds to the probability of a downward jump induced by past upward jumps;
- $\int_0^t \phi_4(t-s)\mathrm{d}N_s^-$ which corresponds to the probability of a downward jump induced by past downward jumps.

Now we will see that, when the ϕ_j have suitable forms the model can reproduce the stylized effects described in the previous section. Moreover, we want to underlying that, due to how the model is built, the price process assumes discrete values, as in the real world.

2.2 Encoding the 2^{nd} property

Since the markets at high frequency are expected to be more efficient then this translate in that, over any period of time, we should have on average the same number of upwards jumps than downwards jumps. This can be translated in:

$$\int_0^t \mathbb{E}[\lambda_s^+]\mathrm{d}s = \mathbb{E}[N_t^+] = \mathbb{E}[N_t^-] = \int_0^t \mathbb{E}[\lambda_s^-]\mathrm{d}s \tag{1}$$

remembering how we have defined λ_t^+ and λ_t^-:

$$\mathbb{E}[\lambda_t^+] = \mu^+ + \int_0^t \phi_1(t-s)\,\mathbb{E}[\lambda_s^+]\mathrm{d}s + \int_0^t \phi_3(t-s)\,\mathbb{E}[\lambda_s^-]\mathrm{d}s$$
$$\mathbb{E}[\lambda_t^-] = \mu^- + \int_0^t \phi_4(t-s)\,\mathbb{E}[\lambda_s^-]\mathrm{d}s + \int_0^t \phi_2(t-s)\,\mathbb{E}[\lambda_s^+]\mathrm{d}s$$

the simplest way to satisfy the equation (1) is to put:

$$\mu^+ = \mu^- \text{ and } \phi_1 + \phi_3 = \phi_2 + \phi_4$$

2.3 Encoding the 3^{rd} property

Market-makers act as liquidity providers, in practice at the beginning they are long inventory, so the ask side is more liquid than the bid side. This translate into the fact that the conditional probability of an upward jump right after an upward jump is smaller than the conditional probability to observe a downward jump after a downward jump. This means that for $t \to 0$ we have:

$$\int_0^t \phi_4(t-s)\mathrm{d}N_s^- > \int_0^t \phi_1(t-s)\mathrm{d}N_s^+$$

or equivalently

$$\int_0^t \phi_2(t-s)\mathrm{d}N_s^+ < \int_0^t \phi_3(t-s)\mathrm{d}N_s^-$$

this can be satisfied in several ways, but we make the strong assumption that exists a constant $\beta > 0$ such that $\phi_3 = \beta\phi_2$. Putting all together we have that the structure of our intensity process is:

$$\begin{pmatrix}\lambda_t^+\\ \lambda_t^-\end{pmatrix} = \mu\begin{pmatrix}1\\1\end{pmatrix} + \int_0^t \begin{pmatrix}\phi_1(t-s) & \beta\phi_2(t-s)\\ \phi_2(t-s) & [\phi_1 + (\beta-1)\phi_2](t-s)\end{pmatrix}\cdot\begin{pmatrix}\mathrm{d}N_s^+\\ \mathrm{d}N_s^-\end{pmatrix}$$

2.4 Encoding the 1^{st} property

Markets have an high degree of endogeneity, which means that the proportion of "non-meaningful" orders with respect to the totality of the orders is close to 1. In order to have an intuition on how to include this effect in our model we need to recall dynamical systems. A dynamical system has a stationary point (or equilibrium) if its spectral radius is less than 1, in the same way we have a kernel transition matrix:

$$\int_0^T \begin{pmatrix}\phi_{1,T}(s) & \beta\phi_{2,T}(s)\\ \phi_{2,T}(s) & [\phi_{1,T} + (\beta-1)\phi_{2,T}](s)\end{pmatrix}\mathrm{d}s = \int_0^t \Phi_T(s)\mathrm{d}s$$

we can extend all the functions on $[0, \infty)$ with the constant zero. We refer to them with a tilde. The spectral radius in our case is equal to:

$$\sigma\left(\int_0^\infty \tilde{\Phi}_T(s)\mathrm{d}s\right) = \|\tilde{\phi}_{1,T}\|_1 + \beta\|\tilde{\phi}_{2,T}\|_1$$

Let $(\Omega, \mathbb{F} = \{(F_T)_{T\geq 0}\}, \mathbb{P})$ a complete filtered probability space. We can find a sequence $\{(\tilde{\phi}_{1,T}; \tilde{\phi}_{2,T})\}_T$ of couple of positive functions each one in the respective $\mathcal{L}^1(F_T)$ such that:

- $\forall T > 0$ we have $\|\tilde{\phi}_{1,T}\|_1 + \beta\|\tilde{\phi}_{2,T}\|_1 < 1$;
- if $T_2 > T_1$ we have both $\tilde{\phi}_{1,T_2} \geq \tilde{\phi}_{1,T_1}$ and $\tilde{\phi}_{2,T_2} \geq \tilde{\phi}_{2,T_1}$;
- satisfying:
$$\lim_{T\to\infty}\left[\|\tilde{\phi}_{1,T}\|_1 + \beta\|\tilde{\phi}_{2,T}\|_1\right] = 1$$

then there exist a limit to this sequence and we will call that $(\tilde{\phi}_1; \tilde{\phi}_2)$ and, due to continuity of the norm, we have that $\|\tilde{\phi}_1\|_1 + \beta\|\tilde{\phi}_2\|_1 = 1$. Then we have built our nearly-unstable system for each $T > 0$ sufficiently large. Moreover notice that, due to the second property of our sequence, we have that the spectral radius, when T is increasing, is also increasing. We will refer to the matrix obtained with $(\tilde{\phi}_1; \tilde{\phi}_2)$ as Φ. The process of the prices up to time $T < \infty$ is now indicated as:

$$P_t^T = N_t^{T,+} - N_t^{T,-}$$

with $N_t^{T,+}$ and $N_t^{T,-}$ with intensity generated by $(\tilde{\phi}_{1,T}; \tilde{\phi}_{2,T})$. From now on, we will denote

$$\sigma\left(\int_0^\infty \tilde{\Phi}_T(s)\mathrm{d}s\right) = a_T = \|\tilde{\phi}_{1,T}\|_1 + \beta\|\tilde{\phi}_{2,T}\|_1$$

with a_T constants and $a_T \uparrow 1$. Moreover, we can also construct the Φ_T as $\Phi = a_T\Phi_T$. We will refer to this as **Assumption 1**.

2.5 Encoding the 4^{th} property

As showed by Jaisson and Rosenbaum, the effect of metaorders are reflected in the Hawkes framework by considering the condition that the kernel matrix exibits heavy-tails. In order to encode the metaorders in the framework we need to put some additional assumptions. Let

$$\Psi_T = \sum_{k\geq 1}(\Phi_T)^{*k}$$

where $(\Phi_T)^{*1} = \Phi_T$ and, for $k > 1$, $(\Phi_T)^{*k}(t) = \int_0^t \Phi_T(s)(\Phi_T)^{*(k-1)}(t-s)\mathrm{d}s$. The **Assumption 2** is that Ψ_T is uniformly bounded, Φ is differentiable and the derivative of each component of Φ is bounded and with finite norm 1. In

order to satisfy this assumption is sufficient, but not necessary, that ϕ_1 and ϕ_2 are both non-increasing functions in $\mathcal{L}^1 \cap \mathcal{L}^\infty$ and differentiable.
The last assumption, **Assumption 3**, is that there exist $\alpha \in (1/2, 1)$ and $C > 0$ such that

$$\alpha t^\alpha \int_t^\infty [\phi_1 + \beta\phi_2](s)\mathrm{d}s \xrightarrow[t\to\infty]{} C$$

and moreover, for some $\mu > 0$ and $\lambda^* > 0$,

$$T^\alpha(1 - a_T) \xrightarrow[t\to\infty]{} \lambda^* \text{ and } T^{1-\alpha}\mu_T \xrightarrow[t\to\infty]{} \mu$$

under this three assumptions also the last stylized effect has been encoded in our model.

2.5.1 Wiener-Hopf equations

Assumption 2 and the definition of the Ψ_T seems a little obscure. This assumption is there only because it allows us to use the following result on integral equations:

Lemma. *Let g be a measurable locally bounded function from $\mathbb{R}$ to $\mathbb{R}^2$ and $\Phi : \mathbb{R}^+ \to \mathcal{M}^2(\mathbb{R})$ be a matrix-value function with integrable components such that the spectral radius of $\int_0^\infty \phi(s)\mathrm{d}s < 1$. Then there exists a unique locally bounded function f from $\mathbb{R}$ to $\mathbb{R}^2$ solution of*

$$f(t) = g(t) + \int_0^t \langle \Phi(t - s), f(s)\rangle \mathrm{d}s, \quad t \geq 0$$

given by

$$f(t) = g(t) + \int_0^t \langle \Psi(t - s), g(s)\rangle \mathrm{d}s, \quad t \geq 0$$

where $\Psi = \sum_{k\geq 1} \Phi^{*k}$.

2.6 From microstructure to macrostructure

If **Assumptions 1,2 and 3** hold then it happens that the asymptotic behaviour of the microstructural model that we have built behaves like an Heston model, more precisely a rough version of it. Indeed, using the same notation as in **Assumption 3**, let:

$$\lambda = \frac{\alpha\lambda^*}{C\Gamma(1 - \alpha)}$$

then it holds true the following theorem:

Theorem. *As $T \to \infty$ then the rescaled microscopic price*

$$\sqrt{\frac{1 - a_T}{\mu T^\alpha}} P_t^T$$

converges in the sense of finite dimensional laws to the following rough Heston model:

$$P_t = \frac{1}{1-(\|\phi_1\|_1 - \|\phi_2\|_1)}\sqrt{\frac{2}{\beta+1}}\int_0^t \sqrt{v_s}\mathrm{d}W_s$$

where v_t is the solution to the following rough SDE:

$$v_t = \frac{1}{\Gamma(\alpha)}\left[\int_0^t (t-s)^{\alpha-1}\lambda(1+\beta-v_s)\mathrm{d}s + \int_0^t (t-s)^{\alpha-1}\lambda\sqrt{\frac{1+\beta^2}{\lambda^*\mu(1+\beta)^2}}\sqrt{v_s}\mathrm{d}\tilde{W}_s\right]$$

where $(W,\tilde{W})$ are two correlated Brownian motions with:

$$\mathrm{d}\langle W,\tilde{W}\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}\mathrm{d}t$$

furthermore, the process v_t has Hölder regularity $\alpha - 1/2 - \varepsilon$ for each $\varepsilon > 0$.

The proof is really technical and out of our scope. It can be found in El Euch, Fukasawa and Rosenbaum.

2.7 Mittag-Leffler functions

As we can see in the previous theorem we are implicitly assuming that $v_0 = 0$. Well, for a practitioner this is a severe limitation. However, it is not immediate to obtain the same result if v_0 is not zero. In this subsection we will give some definitions which we will use to obtain a more general result. First of all, we define the Mittag-Leffler functions. Let $\alpha, \beta \in \mathbb{R}^+$, then the Mittag-Leffler function $E_{\alpha,\beta}$ is defined for $t \in \mathbb{C}$ and $C \in \mathbb{C}$ as

$$E_{\alpha,\beta}(Ct^\alpha) = \sum_{j=0}^{\infty} \frac{C^j t^{j\alpha}}{\Gamma(\alpha j + \beta)}$$

if $\alpha \in (0,1)$ and $\lambda \in \mathbb{R}^+$, we can also define

$$f^{\alpha,\lambda}(t) = \lambda t^{\alpha-1}E_{\alpha,\alpha}(-\lambda t^\alpha)\mathbf{1}_{t\geq 0}$$
$$F^{\alpha,\lambda}(t) = \int_0^t f^{\alpha,\lambda}(s)\mathrm{d}s$$

the first one can be proven to be a density function on $\mathbb{R}^+$ and it is called Mittag-Leffler density function. The Mittag-Leffler density function has many nice properties, among the others, we are interested in the followings:

$$f^{\alpha,\lambda}(t) \sim \frac{\alpha}{\lambda\Gamma(1-\alpha)}t^{-\alpha-1} \quad \text{if} \quad t \to \infty$$
$$F^{\alpha,\lambda}(t) \sim \frac{\lambda}{\Gamma(1+\alpha)}t^{\alpha} \quad \text{if} \quad t \to 0^+$$
$$F^{\alpha,\lambda}(t) \sim 1 + \frac{1}{\lambda\Gamma(1-\alpha)}t^{-\alpha} \quad \text{if} \quad t \to \infty$$
$$D^\alpha[E_\alpha(Ct^\alpha) - 1] = CE_\alpha(Ct^\alpha)$$

when not specified β is assumed to be 1. Moreover we will use this lemma:

Lemma. *Let* $\alpha \in (0,1]$, $u \in \mathbb{C}$ *with* $u = a + ib$ *with* $a \in \mathbb{R}^+$ *and* $b \in [-1/(1-\rho^2), 0]$. *Define as* $C = \sqrt{u(u+i) - \rho^2 u^2}$. *Then for any positive integer* p *and* $t \in \mathbb{R}^+$ *this expansion holds*

$$E_\alpha(-Ct^\alpha) = \sum_{j=1}^{p} \frac{(-1)^{j-1} t^{-j\alpha}}{C^j \Gamma(1-j\alpha)} + \mathcal{O}(|Ct^\alpha|^{-1-p}) \quad \textit{if} \quad t \to \infty$$

2.8 Adjusting the initial volatility

Luckily, we will see that in order to adjust the initial volatility is sufficient to consider an appropriate inhomogeneous intensity for our bi-dimensional Hawkes process and an appropriate kernel matrix. Indeed, using the same notation as in the previous session, we can make a very particular choice for the Φ_T. We will choose the following: suppose that exist $\beta \geq 0$, $\alpha \in (1/2, 1)$ and $\lambda > 0$ such that

$$a_T = 1 - \lambda T^{-\alpha}, \quad \Phi_T(s) = a_T f^{\alpha,1}(s)\chi$$

where

$$\chi = \frac{1}{\beta+1} \begin{pmatrix} 1 & \beta \\ 1 & \beta \end{pmatrix}$$

with this particular choice we see that **Assumption 1, 2** and **3** are satisfied. Moreover, all the first three properties are still well encoded into the model. We need also to notice that χ is an idempotent matrix, this will be useful later when deriving the characteristic function. Now to identify the appropriate inhomogenous intensity $\hat{\mu}_T(\cdot)$ is far more complicated and not so useful, so we will give only the final candidate:

$$\hat{\mu}_T(t) = \mu T^{\alpha-1} + \varepsilon \mu T^{\alpha-1} \left[\frac{1 - \int_0^t a_T f^{\alpha,1}(s) \mathrm{d}s}{1 - a_T} - \int_0^t a_T f^{\alpha,1}(s) \mathrm{d}s \right]$$

with ε, μ positive constants. In the process to obtain this appropriate candidate, written in *The characteristic function of rough Heston models* of Euch and Rosenbaum, they obtained also an explicit form for Ψ_T as

$$\Psi_T(Tt) = \frac{a_T f^{\alpha,\lambda}(t)}{T(1-a_T)}$$

which will be useful in the derivation of the characteristic function. We now need to define the microscopic process converging to the log-price of the rough Heston model, for a positive function $\gamma(\cdot)$, as

$$P_t^T = \sqrt{\frac{\gamma(t)}{2}} \sqrt{\frac{1-a_T}{T^\alpha \mu}} (N_{tT}^{T,+} - N_{tT}^{T,-}) - \frac{\gamma(t)}{2} \frac{1-a_T}{T^\alpha \mu} N_{tT}^{T,+}$$

and finally we have the following

Theorem 1. *As $T \to \infty$, under the assumptions made in this section, the sequence of processes $(P_t^T)_{t\in(0,1)}$ converges in law for the Skorokhod topology to*

$$P_t = \int_0^t \sqrt{v_s}\mathrm{d}\tilde{W}_s - \frac{1}{2}\int_0^t v_s \mathrm{d}s$$

where v is the unique solution of the rough stochastic differential equation

$$v_t = \gamma(t)\varepsilon + \frac{1}{\Gamma(\alpha)}\Bigg[\int_0^t (t-s)^{\alpha-1}\lambda(\gamma(s)-v_s)\mathrm{d}s + \lambda\sqrt{\frac{\gamma(s)(1+\beta^2)}{\lambda\mu(1+\beta)^2}}\int_0^t (t-s)^{\alpha-1}\sqrt{v_s}\mathrm{d}W_s\Bigg]$$

with $(\tilde{W}, W)$ correlated Brownian motions with

$$\mathrm{d}\langle\tilde{W}, W\rangle_t = \frac{1-\beta}{\sqrt{2(1+\beta^2)}}\mathrm{d}t$$

3 The rough Heston model

We have seen that the rough Heston model is what arise from taking the limit of Hawkes processes. Here we add also the drift term to the equation and substitute α with $H + 1/2$. Given a stock price process $S = (S_t)_{t\geq 0}$ the rough Heston model, under risk-free probability measure $\mathbb{Q}$, is the following :

$$\begin{cases} \mathrm{d}S_t = (r-q)S_t\mathrm{d}t + S_t\sqrt{v_t}\mathrm{d}\tilde{W}_t \\ v_t = v_0 + \dfrac{\lambda}{\Gamma(H+\frac{1}{2})}\displaystyle\int_0^t \frac{\gamma(s)-v_s}{(t-s)^{\frac{1}{2}-H}}\mathrm{d}s + \dfrac{\theta}{\Gamma(H+\frac{1}{2})}\displaystyle\int_0^t \frac{\sqrt{v_s}}{(t-s)^{\frac{1}{2}-H}}\mathrm{d}W_s \end{cases}$$

where:

- r is the risk-free rate;
- q is the yield of the underlying;
- W and $\tilde{W}$ are correlated Brownian motions with $\mathrm{d}\langle\tilde{W}, W\rangle_t = \rho\mathrm{d}t$;
- $H \in (0, 1/2)$ is the Hurst exponent of the fractional Brownian motion;
- θ is the volatility of the volatility;
- $\lambda \geq 0$ is a constant representing the "speed" of the mean reversion;
- $\gamma(\cdot)$ is a positive F_0-measurable function representing the mean reversion level for the volatility.

In order to proceed with this chapter we need to define the fractional integral and the fractional derivative. We define the fractional integral of order $\alpha \in (0, 1]$ of a function f as

$$I^\alpha f(t) = \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s) \mathrm{d}s$$

whenever the integral exists. We define the fractional derivative of order $\alpha \in (0, 1]$ of a function f as

$$D^\alpha f(t) = \frac{1}{\Gamma(1-\alpha)} \frac{\mathrm{d}}{\mathrm{d}t} \int_0^t (t-s)^{-\alpha} f(s) \mathrm{d}s$$

whenever it exists.

3.1 Inference of $\lambda\gamma(\cdot)$ from the forward variance curve

In *Perfect hedging in rough Heston models* El Euch and Rosenbaum showed that there is a link between $\lambda\gamma(\cdot)$ and the forward variance curve (intended as in the **Dataset** chapter). We know that the forward variance curve can be obtained by differentiation of the variance swap curve. If the asset have a liquid variance swap market then we have it directly, otherwise we can calculate the fair value of a variance swap using an infinite strip of out-of-the-money options. Assume that the forward variance curve $t \to \mathbb{E}[v_t] =: \xi_0(t)$ admits a fractional derivative of order α then $\lambda\gamma(\cdot)$ can be chosen so that the model is consistent with this market forward variance curve by taking

$$\lambda\gamma(t) = D^\alpha[\xi_0(\cdot) - v_0](t) - \lambda\xi_0(t) \tag{2}$$

Equation (2) can be obtained by calculating $\mathbb{E}[v_t]$ using the second equation of the rough Heston model, showing that the forward variance curve is locally integrable (so the expected value of the stochastic integral is zero), then fractional differentiate the LHS and RHS and reorder. Using this fact and assuming that λ is sufficiently small, we can rewrite the dynamic in a compact way as:

$$(\star) \quad \begin{cases} \mathrm{d}S_t = (r-q)S_t \mathrm{d}t + S_t\sqrt{v_t}\{\rho \mathrm{d}W_t + \sqrt{1-\rho^2}\mathrm{d}W_t^\perp\} \\ v_t = \xi_0(t) + \dfrac{\theta}{\Gamma(H+\frac{1}{2})} \displaystyle\int_0^t \frac{\sqrt{v_s}}{(t-s)^{\frac{1}{2}-H}} \mathrm{d}W_s \end{cases}$$

the hypothesis that λ must be sufficiently small is sensible since the volatility is *slowly* mean reverting.

4 The Characteristic function

Let x_t be the log-spot price, we wish to obtain the characteristic function of terminal log-spot x_T conditional on the initial log-price x_0 and the initial forward variance curve $\xi_0(\cdot)$. In mathematical terms:

$$\phi_{rH}(u, T; 0) = \mathbb{E}_{\mathbb{Q}}\left[e^{iux_T} \middle| x_0, \xi_0(\cdot)\right]$$

4.1 The Characteristic function of an Hawkes process

We need to do a step back and restart from the 2-dimensional Hawkes process presented in the second section of this chapter, called N. Using the same notation as before, let $L(u,t)$ be the characteristic function of the 2-dimensional Hawkes process N conditional at time t:

$$L(u,t) = \mathbb{E}\big[e^{i\langle u, N_t\rangle}\big], \quad u \in \mathbb{R}^2$$

then

Theorem 2. *We have*

$$L(u,t) = \exp\left\{ \int_0^t \langle C(u,t-s) - (1,1), \mu(s)\rangle \mathrm{d}s \right\}$$

where $C : \mathbb{R}^2 \times \mathbb{R}^+ \to \mathbb{C}^2$ is the solution of this integral equation:

$$C(u,t) = \exp\left\{ iu + \int_0^t \Phi_T^\top(s)\big[C(u,t-s) - (1,1)\big]\mathrm{d}s \right\}$$

In order to prove this statement we define two auxiliary independent 2 dimensional point processes $(\tilde{N}_j)_{j=1,2}$. We will refer to $\tilde{N}_1^2$ for the second component of the first auxiliar process, in the same way to the others. Let $(\tilde{N}_j^k)_{j=1,2}$ be two bivariate Hawkes processes with kernel matrix $\Phi : \mathbb{R}^+ \to \mathcal{L}^2(\mathbb{R}^+) \cap \mathcal{L}^1(\mathbb{R}^+)$. We will denote their characteristic function at time t as $L_j(u,t)$. Now, for each j and each $t \geq 0$, let $N_t^{0,1}$ be the number of upwards jumps occurred up to time t and $N_t^{0,2}$ be the number of downwards jumps occurred up to time t, each one is a Poisson process with rates, respectively, $\mu_1(t)$ and $\mu_2(t)$. We also define as $\tau_1^k < \cdots < \tau_{N_t^{0,k}}^k \in [0,t]$ the arrival times of jumps of type k (where $k=1$ for up, $k=2$ for down) of N up to time t. The number of jumps of type k arrived at time τ_u^k has the same law as $(\tilde{N}_{j,t-\tau_u^k}^k)_{j=1,2}$ where $\tilde{N}$ is taken independent from N. So we can write the following equality, in law:

$$N_t^k = N_t^{0,k} + \sum_{j=1}^{2} \sum_{l=1}^{N_t^{0,j}} \tilde{N}_{j,t-\tau_l^j}^{k,l}$$

where $(\tilde{N}_j^{k,l})_{j=1,2}$ are l independent copies of $(\tilde{N}_j^k)_{j=1,2}$, also independent of $(N^{0,k})$. Using these considerations we can obtain:

$$\mathbb{E}\big[e^{i\langle u, N_t\rangle} | N_t^0\big] = e^{i\langle u, N_t^0\rangle} \prod_{j=1}^{2} \prod_{l=1}^{N_t^{0,j}} L_j(u, t-\tau_l^j)$$

Now, fixing k, remember that, conditional on $N^{0,k}, t$, we have that the vector of the arrival times has the same law as the order statistics $(X_{(1)}, \cdots, X_{(N^{0,k},t)})$ built from iid variables $(X_1, \cdots, X_{N_t^{0,k}})$ with density with support in $[0,t]$. We will refer to that density as $\frac{\mu_k(s)\mathbf{1}_{s\leq t}}{\int_0^t \mu_k(s)\mathrm{d}s}$ So, if we use the fact that the X_i are iid we obtain:

$$\mathbb{E}\big[e^{i\langle u,N_t\rangle}|N_t^0\big] = e^{i\langle u,N_t^0\rangle}\prod_{j=1}^{2}\Bigg[\bigg(\int_0^t L_j(u,t-s)\frac{\mu_j(s)}{\int_0^t \mu_j(r)\mathrm{d}r}\mathrm{d}s\bigg)^{N_t^{0,j}}\Bigg]$$

using again independence and rearranging we obtain:

$$L(u,t) = \exp\left\{\sum_{j=1}^{2}\int_0^t \big(e^{iu_j}L_j(u,t-s)-1\big)\mu_j(s)\mathrm{d}s\right\}$$

using the same trick and remembering that also $(\tilde{N}_j^k)_{j=1,2}$ are bivariate Hawkes processes with kernel matrix Φ we can write:

$$L_k(u,t) = \exp\left\{\sum_{j=1}^{2}\int_0^t \big(e^{iu_j}L_j(u,t-s)-1\big)\Phi_{j,k}(s)\mathrm{d}s\right\}$$

now it is enough to define

$$C(u,t) = \begin{pmatrix} e^{iu_1}L_1(u,t) \\ e^{iu_2}L_2(u,t)\end{pmatrix}$$

and do the substitution in the previous equations to obtain the conclusion of the proof.

4.2 Intuition about the result

Using what we obtained in **section 2.8, 3.1** and **4.1** we wish to give an intuition for the following result:

Theorem 3. *Consider the rough Heston model as* $(\star)$ *with* $\rho \in (-1/\sqrt{2}, 1/\sqrt{2}]$. *Then the characteristic function of the terminal log-spot* x_T *conditional on the initial state* (x_0, ξ_0) *is*

$$\phi_{rH}(u,T;0) = \exp\left\{iux_0 + iu(r-q)T + \int_0^T D^{H+1/2}h(u,T-s)\xi_0(s)\mathrm{d}s\right\}$$

where $h(u,\cdot)$ *is the unique continuous solution of the fractional Riccati Cauchy problem*

$$\begin{cases} D^{H+1/2}h(u,\cdot) = -\dfrac{u^2+iu}{2} + iu\theta\rho h(u,\cdot) + \dfrac{\theta^2}{2}h^2(u,\cdot) \\ I^{1/2-H}h(u,0) = 0 \end{cases}$$

NOTE: the assumption that $\rho \in (-1/\sqrt{2}, 1/\sqrt{2}]$ is simply because we have changed parameters

$$\frac{1-\beta}{\sqrt{2(1+\beta^2)}} := \rho$$

with $\beta > 0$ and we need to use **Theorem 1.**
As written in **section 2.8** and remembering the definition of a_T we have that

$$P_t^T = \sqrt{\frac{\lambda\gamma(t)}{2\mu}}T^{-\alpha}(N_{tT}^{T,+} - N_{tT}^{T,-}) - \frac{\lambda\gamma(t)}{2\mu}T^{-2\alpha}N_{tT}^{T,+}$$

and we know that, if $T \to \infty$, this sequence of processes converges in law to P where $P_t = \log(S_t/S_0)$. Now let $N^T = (N^{T,+}, N^{T,-})$ be a sequence of two dimensional Hawkes processes (varying T) satisfying all the assumptions in **section 2.8** and denote with $L^T(u,t)$ the characteristic function of the process N^T at point $u = (u^+, u^-) \in \mathbb{C}^2$ and time t. If we fix a scalar $\overline{u} \in \mathbb{C}$ and let it be

$$\overline{u}_+^T = \overline{u}\sqrt{\frac{\lambda\gamma(tT)}{2\mu}}T^{-\alpha} - \overline{u}\frac{\lambda\gamma(tT)}{2\mu}T^{-2\alpha}, \quad \overline{u}_-^T = -\overline{u}\sqrt{\frac{\lambda\gamma(tT)}{2\mu}}T^{-\alpha}$$

then we have that, since convergence in law implies pointwise convergence of the characteristics we have that

$$L^T((\overline{u}_+^T, \overline{u}_-^T), tT) \to L(\overline{u}, t) \quad \text{if} \quad T \to \infty \tag{3}$$

where $L(\cdot, t)$ is the characteristic function of P at time t. Now, we have to notice the following fact, which will be useful later: taking the definition of $\hat{\mu}(\cdot)$ given in **section 2.8** and the asyntotic properties of the Mittag-Leffler density function given in **section 2.7** we may write, for each $t \in (0, 1]$, that

$$\begin{aligned}
T^{1-\alpha}\hat{\mu}(tT) &= T^{1-\alpha}\mu_T + \varepsilon T^{1-\alpha}\mu_T\left[\frac{T^\alpha}{\lambda}\int_{tT}^{\infty} f^{\alpha,1}(s)\mathrm{d}s + \lambda T^{-\alpha}\int_0^{tT} f^{\alpha,1}(s)\mathrm{d}s\right] \\
&= \mu_T\left[1 + \frac{\varepsilon t^{-\alpha}}{\lambda}\cdot(tT)^\alpha\int_{tT}^{\infty} f^{\alpha,1}(s)\mathrm{d}s\right] + \mu_T\varepsilon\lambda T^{-\alpha}\int_0^{tT} f^{\alpha,1}(s)\mathrm{d}s\Big] \\
&\xrightarrow[T\to\infty]{} \mu + \frac{\mu\varepsilon t^{-\alpha}}{\lambda\Gamma(1-\alpha)}
\end{aligned}$$

Thanks to **Theorem 2.** we can write the characteristic function of our Hawkes process as

$$\begin{aligned}
L^T((\overline{u}_+^T, \overline{u}_-^T), tT) = \exp\Bigg\{\int_0^{tT} &\hat{\mu}_T(s)\big[C^{T,+}((\overline{u}_+^T, \overline{u}_-^T), tT - s) - 1) \\
&+ C^{T,-}((\overline{u}_+^T, \overline{u}_-^T), tT - s) - 1)\big]\mathrm{d}s\Bigg\}
\end{aligned}$$

where $C^T((\overline{u}_+^T, \overline{u}_-^T), t) = \big(C^{T,+}((\overline{u}_+^T, \overline{u}_-^T), t), C^{T,-}((\overline{u}_+^T, \overline{u}_-^T), t)\big)$ is the solution to its respective integral equation written in **Theorem 2.**. Now we define

$$Y^T(\overline{u}, t) = \big(Y^{T,+}(\overline{u}, t), Y^{T,-}(\overline{u}, t)\big) = C^T((\overline{u}_+^T, \overline{u}_-^T), tT)$$

and we can rewrite the characteristic function as

$$L^T((\overline{u}_+^T,\overline{u}_-^T),tT) = \exp\left\{ \int_0^t T^\alpha \big[\big(Y^{T,+}(\overline{u},t-s)-1\big) + \big(Y^{T,-}(\overline{u},t-s)-1\big)\big] \cdot \big[T^{1-\alpha}\hat{\mu}(sT)\big]\mathrm{d}s \right\}$$

Since we have (3) we can expect that as $T \to \infty$ then $T^\alpha\big(Y^T(\overline{u},t)-(1,1)\big)$ converges to some functions $\big(c(\overline{u},t),d(\overline{u},t)\big)$, this can be shown as in the last section of *The characteristic function of rough Heston models* of El Euch and Rosenbaum. Using the fact that $\big(Y^T(\overline{u},t)-(1,1)\big) = \mathcal{O}(T^{-\alpha})$ (in the sense component-wise) we can expand $\log\big(Y^T(\overline{u},t)\big)$ around $(1,1)$ (where $\log(\cdot)$ has been applied on each component) and obtain

$$\log\big(Y^T(\overline{u},t)\big) = Y^T(\overline{u},t)-(1,1) - \frac{1}{2}\big(Y^T(\overline{u},t)-(1,1)\big)^2 + o(T^{-2\alpha})(t)$$

and using the characteristic function above and solving for $Y^T(\overline{u},t)-(1,1)$ we obtain

$$\begin{aligned} Y^T(\overline{u},t)-(1,1) = {} & i\overline{u}\sqrt{\frac{\lambda\gamma(t)}{2\mu}}(1,-1)T^{-\alpha} - i\overline{u}\frac{\lambda\gamma(t)}{2\mu}(1,0)T^{-2\alpha} \\ & + T\int_0^t \phi_T^\top(Ts)\big(Y^T(\overline{u},t-s)-(1,1)\big)\mathrm{d}s \\ & + \frac{1}{2}\big(Y^T(\overline{u},t)-(1,1)\big)^2 + o(T^{-2\alpha})(t) \end{aligned}$$

where with $\big(Y^T(\overline{u},t)-(1,1)\big)^2$ we intend $\langle\big(Y^T(\overline{u},t)-(1,1)\big),\big(Y^T(\overline{u},t)-(1,1)\big)\rangle$. Now, we wish to use the lemma in **section 2.5.1** to solve the integral part and then using the explicit form for $\Psi_T(T\cdot)$ written in **section 2.8**. So we need to notice

$$\begin{aligned} \sum_{k\geq 1}\big(T\Phi_T(T\cdot)\big)^{*k} &= \sum_{k\geq 1} T\Phi_T^{*k}(T\cdot) \\ &= T\sum_{k\geq 1}(a_T f^{\alpha,\lambda})^{*k}(T\cdot)\chi^k \\ &= T\Psi_T(T\cdot)\chi \\ &= \frac{a_T T^\alpha}{\lambda(\beta+1)} f^{\alpha,\lambda}(\cdot)\begin{pmatrix} 1 & \beta \\ 1 & \beta \end{pmatrix} \end{aligned}$$

and then, applying the lemma, we obtain

$$
\begin{aligned}
Y^T(\overline{u},t)-(1,1) = {} & i\overline{u}\sqrt{\frac{\lambda\gamma(t)}{2\mu}}(1,-1)T^{-\alpha} - i\overline{u}\frac{a_T\gamma(t)}{2\mu(1+\beta)}(1,\beta)T^{-\alpha}F^{\alpha,\lambda}(t) \\
& + \frac{a_T T^{\alpha}}{2\lambda(\beta+1)}\int_0^t f^{\alpha,\lambda}(s)\begin{pmatrix}1 & 1\\ \beta & \beta\end{pmatrix}\big(Y^T(\overline{u},t)-(1,1)\big)^2\mathrm{d}s \\
& + o(T^{-\alpha})(t)
\end{aligned}
$$

so we have that $\big(c(\overline{u},t),d(\overline{u},t)\big)$ must satisfy the integral equations

$$
\begin{aligned}
c(\overline{u},t) &= i\overline{u}\sqrt{\frac{\lambda\gamma(t)}{2\mu}} - i\overline{u}\frac{\gamma(t)}{2\mu(1+\beta)}F^{\alpha,\lambda}(t) \\
&\quad + \frac{1}{2\lambda(\beta+1)}\int_0^t f^{\alpha,\lambda}(s)\big(c^2(\overline{u},t-s)+d^2(\overline{u},t-s)\big)\mathrm{d}s \\
d(\overline{u},t) &= -i\overline{u}\sqrt{\frac{\lambda\gamma(t)}{2\mu}} - i\overline{u}\frac{\beta\gamma(t)}{2\mu(1+\beta)}F^{\alpha,\lambda}(t) \\
&\quad + \frac{\beta}{2\lambda(\beta+1)}\int_0^t f^{\alpha,\lambda}(s)\big(c^2(\overline{u},t-s)+d^2(\overline{u},t-s)\big)\mathrm{d}s
\end{aligned}
$$

and defining $h(\overline{u},t) \coloneqq \mu[c(\overline{u},t)+d(\overline{u},t)]$ then we have that

$$
\begin{aligned}
L(\overline{u},t) &= \exp\left\{\lambda\gamma(t)I^1h(\overline{u},t)+\varepsilon\gamma(t)I^{1-\alpha}h(\overline{u},t)\right\} \\
&= \exp\left\{\lambda\gamma(t)I^1h(\overline{u},t)+v_0I^{1-\alpha}h(\overline{u},t)\right\}
\end{aligned}
$$

where h is the solution of the fractional Riccati Cauchy problem

$$
\begin{cases}
D^{\alpha}h(u,t) = -\dfrac{u^2+iu}{2}+\lambda(iu\theta\rho-1)h(u,t)+\dfrac{\lambda^2\theta^2}{2}h^2(u,t) \\
I^{1-\alpha}h(u,0)=0
\end{cases}
$$

now if we reformulate in terms of the forward variance curve, add a drift term and suppose that the starting log-spot price is not 1 we obtain the result that we wanted to prove at the beginning of this section.

4.3 Rational approximation of the solution

We have a quasi-closed form for the characteristic function and we wish to obtain the solution to the fractional equation. Unfortunately, this solution is not known, so we will use the Padè approximants to obtain a fast and reliable approximation. In this section we will follow the work of Gatheral and Radoicic in *Rational approximation of the rough Heston solution.* In particular we will derive an expansion for small times for the characteristic function, an expansion for long times and then we will derive a Padé rational expansion to match the two formulae.

4.3.1 Small times expansion

For the small times expansion we will follow the work of *Alòs et al.*. Only for this paragraph suppose that interest rates, borrow costs and yields are zero. Let $H\big(x_t, w_t(T)\big)$ be a solution of this equation

$$-\frac{\partial H}{\partial w}\big(x_t, w_t(T)\big) + \frac{1}{2}\frac{\partial^2 H}{\partial x^2}\big(x_t, w_t(T)\big) - \frac{1}{2}\frac{\partial H}{\partial x}\big(x_t, w_t(T)\big) = 0$$

where

$$w_t(T) = \mathbb{E}\bigg[\int_t^T v_s \mathrm{d}s \bigg| F_t\bigg] = \int_0^T \xi_0(s)\mathrm{d}s - \int_0^t v_s \mathrm{d}s =: M_t - \int_0^t v_s \mathrm{d}s$$

and x_t is the log price as before. Now we need these definition

Definition 1. *Let A_t and B_t two stochastic processes. Then*

$$(A \diamond B)_t(T) := \mathbb{E}\bigg[\int_t^T d\langle A_t, B_t\rangle_s \bigg| F_t\bigg]$$

provided that the expectation is finite.

Definition 2. *Let $H_t := H\big(x_t, w_t(T)\big)$, defined as before, then*

$$(x \diamond M)_t(T) \cdot H_t := \mathbb{E}\bigg[\int_t^T d\langle x_t, M_t\rangle_s \bigg| F_t\bigg]\frac{\partial^2 H_t}{\partial x \partial w}$$

Definition 3. *Let $\mathbb{F}_0 = M$. Then the forest of order $k \in \mathbb{N}$ is defined recursively as:*

$$\mathbb{F}_k = \frac{1}{2}\sum_{l=0}^{k-2}\sum_{j=0}^{k-2} \mathbf{1}_{l+j=k-2}\mathbb{F}_l \diamond \mathbb{F}_j + x \diamond \mathbb{F}_{k-1}$$

In the last definition we dropped the subscrip and the point of evaluation for $\diamond$, from now on, unless specified, it is always T. Using this notation Alòs, Gatheral and Radoicic obtained this powerful result

Theorem 4. *If H_t is a solution of the differential equation presented at the beginning of this section, $\mathbb{E}[H_T|F_t]$ is finite and for each $j \geq 0$ the integrals in each forest $\mathbb{F}_j$ exist. Then*

$$\mathbb{E}[H_T|F_t] = e^{\sum_{j=1}^{\infty}\mathbb{F}_j} \cdot H_t$$

where the exponential is to be understood as a formal power series and $\cdot$ is the operator in ***Definition 2.***

this theorem gives us an exact representation of the conditional expectation for every model that can be written in the forward variance form without assuming Markovianity. We can now rewrite the rough Heston model in the forward variance form as:

$$
\begin{cases}
\mathrm{d}S_t = S_t\sqrt{v_t}\{\rho \mathrm{d}W_t + \sqrt{1-\rho^2}\mathrm{d}W_t^{\perp}\} =: S_t\sqrt{v_t}\mathrm{d}\tilde{W}_t \\
\mathrm{d}\xi_t(u) = \dfrac{\theta}{\Gamma(H+\frac{1}{2})}\dfrac{\sqrt{v_t}}{(u-t)^{\frac{1}{2}-H}}\mathrm{d}W_t
\end{cases}
$$

and, excluding the term which are F_t-measurable and do not contribute in the tree computations

$$
\begin{aligned}
\mathrm{d}x_t &= \sqrt{v_t}\mathrm{d}\tilde{W}_t + F_t\text{-measurable terms} \\
\mathrm{d}M_t &= \frac{\theta(T-t)^{\alpha}}{\Gamma(\alpha+1)}\sqrt{v_t}\mathrm{d}W_t
\end{aligned}
$$

and proceeding with the computations

$$
\mathbb{F}_1 = x \diamond M = \frac{\rho\theta}{\Gamma(\alpha+1)}\int_t^T \xi_t(s)(T-s)^{\alpha}\mathrm{d}s
$$

if we define for $j \in \mathbb{N}$

$$
I_t^{(j)}(T) \coloneqq \int_t^T \xi_t(s)(T-s)^{j\alpha}\mathrm{d}s
$$

then we have

$$
\begin{aligned}
\mathrm{d}I_t^{(j)}(T) &= \int_t^T (T-s)^{j\alpha}\mathrm{d}\xi_t(s)\mathrm{d}s \\
&= \frac{\theta\Gamma(1+j\alpha)}{\Gamma(1+j\alpha+\alpha)}\sqrt{v_t}(T-t)^{(j+1)\alpha}\mathrm{d}W_t + \text{ drift terms}
\end{aligned}
$$

and in the computation of $\diamond$ the drift terms do not contribute. With this notation we have:

$$
x \diamond M = \frac{\rho\theta}{\Gamma(\alpha+1)}I_t^{(1)}(T)
$$

in $\mathbb{F}_2$ we have two trees:

$$
\begin{aligned}
M \diamond M &= \frac{\theta^2}{\Gamma(\alpha+1)^2}I_t^{(2)}(T) \\
x \diamond (x \diamond M) &= \frac{\rho\theta}{\Gamma(\alpha+1)}\mathbb{E}\left[\int_t^T \mathrm{d}\langle x, I^{(1)}\rangle_s \mathrm{d}s \middle| F_t\right] \\
&= \frac{\rho^2\theta^2}{\Gamma(2\alpha+1)}I_t^{(2)}(T)
\end{aligned}
$$

it can be proven by induction that each tree in the forest $\mathbb{F}_j$ is equal to $\theta^j I_t^{(j)}(T)$ multiplied by a constant. Now, let's consider this characteristic function

$$H_t(u) = \phi(u,T;t) \coloneqq \exp\left\{iux_t - \frac{u^2+ui}{2}w_t(T)\right\}$$

this clearly satisfy the PDE at the beginning of this section and, moreover, it holds true, trough differentiation, that

$$e^{\sum_{j=1}^{\infty}\mathbb{F}_j} \cdot \phi(u,T;t) = e^{\sum_{j=1}^{\infty}\tilde{\mathbb{F}}_j(u)}\phi(u,T;t)$$

where $\tilde{\mathbb{F}}_j(u)$ is defined as $\mathbb{F}_j$ but with each occurrence of $\partial/\partial w$ replaced with $-(u^2+ui)/2$ and each occurrence of $\partial/\partial x$ replaced with iu. Using **Theorem 4.** we have that

$$\phi_{rH}(u,T;t) = \mathbb{E}_{\mathbb{Q}}\left[e^{iux_T}\middle|F_t\right] = e^{\sum_{j=1}^{\infty}\tilde{\mathbb{F}}_j(u)}\phi(u,T;t)$$

since we have used **Theorem 4.** we have to be sure that each tree in each forest $\mathbb{F}_j$ exists, this is true if the forward variance curve is bounded on finite intervals, in that case each tree is of order $(T-t)^{j\alpha+1}$. From the calculations that we made before on $\mathbb{F}_k$ we have that

$$\tilde{\mathbb{F}}_k(u) = \beta_k(u)\theta^k I_t^{(k)}(T)$$

with $\beta_k(u)$ a coefficient dependent on k and u. Define also $\tilde{X}(u) = iux_t$. Firstly we compute for $l<j$

$$\begin{aligned}
\tilde{\mathbb{F}}_l(u) \diamond \tilde{\mathbb{F}}_j(u) &= \theta^{l+j+2}\beta_l(u)\beta_j(u)\frac{\Gamma(l\alpha+1)\Gamma(j\alpha+1)}{\Gamma(l\alpha+\alpha+1)\Gamma(j\alpha+\alpha+1)}I_t^{(l+j+2)}(T) \\
\tilde{X}(u) \diamond \tilde{\mathbb{F}}_k(u) &= iu\rho\theta^k\frac{\Gamma(k\alpha-\alpha+1)}{\Gamma(k\alpha+1)}\beta_{k-1}(u)
\end{aligned}$$

Using the recursion formula in **Definition 3.** with the correct modifications (substituting $\mathbb{F}_j$ with $\tilde{\mathbb{F}}_j(u)$ and x with $\tilde{X}(u)$) we obtain a recursion formula to express the coefficients $\beta_j(u)$ as

$$\begin{aligned}
\beta_0(u) &= -\frac{u^2+iu}{2} \\
\beta_k(u) &= \frac{1}{2}\sum_{l=0}^{k-2}\sum_{j=0}^{k-2}\mathbf{1}_{l+j=k-2}\beta_i(u)\beta_j(u)\frac{\Gamma(l\alpha+1)\Gamma(j\alpha+1)}{\Gamma(l\alpha+\alpha+1)\Gamma(j\alpha+\alpha+1)} \\
&\quad + iu\rho\frac{\Gamma(k\alpha-\alpha+1)}{\Gamma(k\alpha+1)}\beta_{k-1}(u)
\end{aligned}$$

and now defining $h(u,t)$ as the formal power series

$$h(u,t) = \sum_{j=0}^{\infty}\frac{\Gamma(\alpha j+1)}{\Gamma(\alpha j+\alpha+1)}\beta_j(u)\theta^j t^{(j+1)\alpha}$$

we have that for small times is converging, satisfy the fractional Riccati equation, the boundary condition and, moreover, thanks to what we noticed before

$$\phi_{rH}(u,T;t) = e^{\sum_{j=1}^{\infty}\tilde{\mathbb{F}}_j(u)}\phi(u,T;t) = \exp\left\{iux_t + \int_t^T D^{\alpha}h(u,T-s)\xi_t(s)\mathrm{d}s\right\}$$

which was exactly what we obtained in the previous sections (minus the drift term). Notice that the function $\theta h(u,t)$ depends only on the quantity θt^{α}, so if we do the following change of variable $\tilde{t}^{\alpha} = \theta t^{\alpha}$ we can rewrite our fractional differential equation as

$$\begin{aligned} D^{\alpha}h(u,\tilde{t}) &= -\frac{u^2+ui}{2} + iu\rho h(u,\tilde{t}) + \frac{1}{2}h(u,\tilde{t})^2 \\ &= \frac{1}{2}\left[h(u,\tilde{t}) - r_-\right]\left[h(u,\tilde{t}) - r_+\right] \end{aligned}$$

with $r_{\pm} = -iu\rho \pm \sqrt{u^2 + iu - \rho^2 u^2}$.

4.3.2 Long times expansion

If we define $C \coloneqq (r_+ - r_-)/2$. Then this proposition holds:

Proposition 5. *Let $h_{\infty}(u,\tilde{t}) \coloneqq r_-[1 - E_{\alpha}(-C\tilde{t}^{\alpha})]$. For $u \in \mathbb{C}$ with $\Re(u) \geq 0$ and $\tilde{t} \in \mathbb{R}^+$, if $\tilde{t} \to \infty$ then $h_{\infty}(u,\tilde{t})$ solves the fractional equation in **Theorem 3.** up to an error term of $\mathcal{O}(|C\tilde{t}^{\alpha}|^{-2})$.*

To prove this proposition is sufficient to apply the last property shown in **section 2.7** and use the lemma. Notice also that the definition of C in the lemma is coherent with the definition of C in this section. Looking at **Proposition 4.** it raises a natural *ansatz* for $h(u,\tilde{t})$ when $\tilde{t} \to \infty$ and it is

$$h(u,\tilde{t}) = r_- \sum_{j=0}^{\infty} \gamma_j \frac{\tilde{t}^{-j\alpha}}{C^j\Gamma(1-j\alpha)}$$

for some coefficients $(\gamma_j)_{j=0}^{\infty}$. Using now the last property of Mittag-Leffler functions we, after changing the index, obtain

$$\frac{1}{r_-}D^{\alpha}h(u,\tilde{t}) = C\sum_{j=1}^{\infty}\gamma_{j-1}\frac{\tilde{t}^{-j\alpha}}{C^j\Gamma(1-j\alpha)}$$

then, assuming that our *ansatz* is the solution to the fractional differential equation we have also the following

$$\begin{aligned} \frac{1}{r_-}D^{\alpha}h(u,\tilde{t}) &= \frac{1}{r_-}\frac{1}{2}(h(u,\tilde{t}) - r_-)(h(u,\tilde{t}) - r_+) \\ &= \sum_{j=1}^{\infty}\gamma_j\frac{\tilde{t}^{-j\alpha}}{C^j\Gamma(1-j\alpha)}\Bigg(-C + \frac{r_-}{2}\sum_{j=1}^{\infty}\gamma_j\frac{\tilde{t}^{-j\alpha}}{C^j\Gamma(1-j\alpha)}\Bigg) \end{aligned}$$

using the identity principle for power series we obtain

$$
\begin{aligned}
\gamma_0 &= 1 \\
\gamma_1 &= -1 \\
\gamma_2 &= 1 + \frac{r_-}{2C}\frac{\Gamma(1-2\alpha)}{\Gamma(1-\alpha)^2} \\
&\vdots \\
\gamma_j &= -\gamma_{j-1} + \frac{r_-}{2C}\sum_{i=1}^{\infty}\sum_{l=1}^{\infty}\mathbf{1}_{i+l=j}\gamma_i\gamma_l\frac{\Gamma(1-j\alpha)}{\Gamma(1-i\alpha)\Gamma(1-l\alpha)}
\end{aligned}
$$

4.3.3 Padé approximation

We define the rational approximation of $h(u,\tilde{t})$ with m terms at the numerator and n terms at the denominator as

$$h^{(m,n)}(u,\tilde{t}) = \frac{\sum_{j=1}^{m} p_j \tilde{t}^{j\alpha}}{\sum_{j=1}^{n} q_j \tilde{t}^{j\alpha}}$$

such that

$$h^{(m,n)}(u,\tilde{t})\sum_{j=1}^{n} q_j \tilde{t}^{j\alpha} - \sum_{j=1}^{m} p_j \tilde{t}^{j\alpha} = \mathcal{O}\big(\tilde{t}^{\alpha(m+n+1)}\big)$$

Notice that for $t \to \infty$, thanks to the expansion for long times, we have that

$$h^{(m,n)}(u,\tilde{t}) \sim \frac{p_m \tilde{t}^{m\alpha}}{q_n \tilde{t}^{n\alpha}}$$

and we have seen that for long times it must be finite, so this can happen if and only if $m = n$. So the only admissible approximation of h are the ones of type $h^{(m,m)}$. As denoted in the paper we will choose $m = 3$, although there is no theoretical motivation for this choice, in practice happens to be both fast to compute and sufficiently accurate for our scope. Then, setting WLOG the constant at the denominator equal to 1:

$$h^{(3,3)}(u,\tilde{t}) = \frac{p_1\tilde{t}^{\alpha} + p_2\tilde{t}^{2\alpha} + p_3\tilde{t}^{3\alpha}}{1 + q_1\tilde{t}^{\alpha} + q_2\tilde{t}^{2\alpha} + q_3\tilde{t}^{3\alpha}}$$

thanks to the small time expansion for small x we have, for some coefficients b_1, b_2, b_3:

$$h(u,\tilde{t}) = b_1\tilde{t}^{\alpha} + b_2\tilde{t}^{2\alpha} + b_3\tilde{t}^{3\alpha} + \mathcal{O}(\tilde{t}^{4\alpha})$$

meanwhile for long times and coefficients c_0, c_1, c_2:

$$h(u,\tilde{t}) = c_0 + \frac{c_1}{\tilde{t}^{\alpha}} + \frac{c_2}{\tilde{t}^{2\alpha}} + \mathcal{O}\big(\tilde{t}^{-3\alpha}\big)$$

using the definition at the beginning of this section we obtain the following equations

$$
\begin{aligned}
p_1 &= b_1 \\
p_2 - p_1 q_1 &= b_2 \\
p_1 q_1^2 - p_1 q_2 - p_2 q_1 + p_3 &= b_3 \\
p_3 &= c_0 q_3 \\
p_2 q_3 - p_3 q_2 &= c_1 q_3^2 \\
p_1 q_3^2 - p_2 q_2 q_3 - p_3 q_1 q_3 + p_3 q_2^2 &= c_2 q_3^3
\end{aligned}
$$

this linear system can be solved and the solution is

$$
\begin{aligned}
p_1 &= b_1 \\
p_2 &= \frac{b_1^3 c_1 + b_1^2 c_0^2 + b_1 b_2 c_0 c_1 - b_1 b_3 c_0 c_2 + b_1 b_3 c_1^2 + b_2^2 c_0 c_2 - b_2^2 c_1^2 + b_2 c_0^3}{b_1^2 c_2 + 2 b_1 c_0 c_1 + b_2 c_0 c_2 - b_2 c_1^2 + c_0^3} \\
p_3 &= c_0 q_3 \\
q_1 &= \frac{b_1^2 c_1 - b_1 b_2 c_2 + b_1 c_0^2 - b_2 c_0 c_1 - b_3 c_0 c_2 + b_3 c_1^2}{b_1^2 c_2 + 2 b_1 c_0 c_1 + b_2 c_0 c_2 - b_2^2 c_1^2 + c_0^3} \\
q_2 &= \frac{b_1^2 c_0 - b_1 b_2 c_1 - b_1 b_3 c_2 + b_2^2 c_2 + b_2 c_0^2 - b_3 c_0 c_1}{b_1^2 c_2 + 2 b_1 c_0 c_1 + b_2 c_0 c_2 - b_2^2 c_1^2 + c_0^3} \\
q_3 &= \frac{b_1^3 + 2 b_1 b_2 c_0 + b_1 b_3 c_1 - b_2^2 c_1 + b_3 c_0^2}{b_1^2 c_2 + 2 b_1 c_0 c_1 + b_2 c_0 c_2 - b_2^2 c_1^2 + c_0^3}
\end{aligned}
$$

notice that $I^{1-\alpha} h^{(3,3)}(u,0) = 0$.

5 Pricing

Using the facts that h is a solution of the fractional Riccati equation and that $h^{(3,3)}$ approximates it well then we have:

$$
\begin{aligned}
\phi_{rH}(u,T;0) = \exp\Bigg\{ & iux_0 + iu(r-q)T - \int_0^T \frac{u^2 + iu}{2} \xi_0(s) \mathrm{d}s \\
& + \int_0^T iu\theta\rho h(u, T-s) \xi_0(s) \mathrm{d}s \\
& + \int_0^T \frac{\theta^2}{2} h^2(u, T-s) \xi_0(s) \mathrm{d}s \Bigg\}
\end{aligned}
$$

and then

$$
\begin{aligned}
\hat{\phi}_{rH}(u,T;0) = \exp\Bigg\{ & iux_0 + iu(r-q)T - \int_0^T \frac{u^2+iu}{2}\xi_0(s)\mathrm{d}s \\
& + \int_0^T iu\theta\rho h^{(3,3)}(u,T-s)\xi_0(s)\mathrm{d}s \\
& + \int_0^T \frac{\theta^2}{2}\big[h^{(3,3)}(u,T-s)\big]^2\xi_0(s)\mathrm{d}s \Bigg\}
\end{aligned}
$$

where $\hat{\phi}_{rH}(u,T;0) \approx \phi_{rH}(u,T;0)$. Now if we denote the volatility of a variance swap at time 0 with tenor T with $\hat{\sigma}_0^T$ and remember the definition of the forward variance curve we have:

$$
T\hat{\sigma}_0^T = \int_0^T \xi_0(s)\mathrm{d}s
$$

where we have to notice also that $\left(\hat{\sigma}_0^T\right)^2$ is the strike of a variance swap, which, for some tenor, can be observed in the market. The parametrization for the volatility of a variance swap is the Gompertz function which is the following

$$
\hat{\sigma}_0^T = z_1 e^{-z_2 e^{-z_3 T}}
$$

where z_1, z_2 and z_3 are positive constants. The effects of the parameters are: z_1 is the asymptote (i.e. the long time implied future volatility), z_2 sets the displacement along the x-axis (i.e. time to maturity) and z_3 sets the growth rate. This three parameters are then fitted using least-squares to the observable prices of variance swaps. Now our formula becomes

$$
\begin{aligned}
\hat{\phi}_{rH}(u,T;0) = \exp\Bigg\{ & iux_0 + iuT\left[r - q + (iu-1)\frac{\hat{\sigma}_0^T}{2}\right] \\
& + iu\theta\rho \int_0^T h^{(3,3)}(u,T-s)\xi_0(s)\mathrm{d}s \\
& + \frac{\theta^2}{2}\int_0^T \big[h^{(3,3)}(u,T-s)\big]^2\xi_0(s)\mathrm{d}s \Bigg\}
\end{aligned}
$$

In the Heston chapter we have shown the COS method for pricing European claims, in particular European vanilla puts. Since this method is really general we wish to adapt it to price European puts under the rough Heston model. Let K be the strike, T be the tenor and $P_{0,K}$ the price of the put at time 0 then, using the same notation as in the Heston chapter, we have

$$
\begin{aligned}
P_{0,K} & \approx e^{-rT}\left[\frac{1}{2}\Re\big\{\phi_{rH}(0,T;0)\big\}V_0 + \sum_{k=1}^{N-1}\Re\left\{\phi_{rH}\left(\frac{k\pi}{b-a},T;0\right)\cdot e^{-\frac{ik\pi a}{b-a}}\right\}V_k\right] \\
& \approx e^{-rT}\left[\frac{1}{2}\Re\big\{\hat{\phi}_{rH}(0,T;0)\big\}V_0 + \sum_{k=1}^{N-1}\Re\left\{\hat{\phi}_{rH}\left(\frac{k\pi}{b-a},T;0\right)\cdot e^{-\frac{ik\pi a}{b-a}}\right\}V_k\right]
\end{aligned}
$$

where, using Le Floc'h's correction, we have

$$V_k = \frac{2}{b-a}\big[K\psi_k(a,-x) - S_0\chi_k(a,-x)\big]$$

and

$$\psi_k(a,-x) = \begin{cases} \dfrac{a-b}{k\pi}\sin\left(k\pi\dfrac{x+a}{b-a}\right) & k \neq 0 \\ -x-a & k=0 \end{cases}$$

$$\chi_k(a,-x) = \frac{1}{1+\left(\frac{k\pi}{b-a}\right)^2}\left[e^{-x}\cos\left(k\pi\frac{x+a}{b-a}\right) - e^a - \frac{k\pi}{b-a}e^{-x}\sin\left(k\pi\frac{x+a}{b-a}\right)\right]$$

the only thing which is missing is how to choose the truncation interval $[a,b]$.

5.1 Computation of a and b

As in the other chapter we will use the formula proposed by Fang & Osterlee to compute the truncation range as:

$$[a,b] = [c_1 - 12\sqrt{|c_2|}, c_1 + 12\sqrt{|c_2|}]$$

where c_1 and c_2 are the two first cumulants. Suppose that $x_0 = 0$ (we are pricing an ATM option). The cumulant generating function of the rough Heston model can be written as

$$\begin{aligned}\psi(u,T;0) &= uT(r-q) + \int_0^T D^\alpha h(-iu,T-s)\xi_0(s)\mathrm{d}s \\ &= uT\left[r-q+(u-1)\frac{\hat{\sigma}_0^T}{2}\right] \\ &+ u\theta\rho\int_0^T h(-iu,T-s)\xi_0(s)\mathrm{d}s \\ &+ \frac{\theta^2}{2}\int_0^T \big[h(-iu,T-s)\big]^2\xi_0(s)\mathrm{d}s\end{aligned}$$

so we have

$$\begin{aligned}c_1 = \psi'(0,T;0) &= T(r-q) - i\int_0^T D^\alpha h'(0,T-s)\xi_0(s)\mathrm{d}s \\ &= T(r-q) - \frac{1}{2}\int_0^T \xi_0(s)\mathrm{d}s \\ &= T\left(r-q-\frac{\hat{\sigma}_0^T}{2}\right)\end{aligned}$$

and for the second cumulant

$$
\begin{aligned}
c_2 = \psi''(0,T;0) &= 2\theta\rho\int_0^T h'(0,T-s)\xi_0(s)\mathrm{d}s + \theta^2\int_0^T \big[h'(0,T-s)\big]^2\xi_0(s)\mathrm{d}s \\
&= -\frac{\theta\rho}{\Gamma(H+3/2)}\int_0^T (T-s)^{H+1/2}\xi_0(s)\mathrm{d}s \\
&\quad + \frac{\theta^2}{4[\Gamma(H+3/2)]^2}\int_0^T (T-s)^{2H+1}\xi_0(s)\mathrm{d}s \\
&\approx \frac{\theta\rho z_1^2 T^{H+3/2}}{(H+3/2)\Gamma(H+3/2)} - \frac{\theta^2 z_1^2 T^{2H+2}}{4(2H+2)[\Gamma(H+3/2)]^2}
\end{aligned}
$$

where in the last step we have used the fact that the variance swaps are parametrized using a Gompertz function with asymptote z_1.

6 Calibration

The calibration is done in two steps. Firstly, we use a least square algorithm to fit the prices of the variance swaps to the ones observed in the market. In this way we fix z_1, z_2 and z_3. After that, let $\Xi := [H, \rho, \theta]^\top$ be our vector of parameters. We denote with $\sigma^*(K_i, T_i)$ the market implied volatility for calls with strike K_i and maturity T_i and with $\sigma(\Xi; K_i, T_i)$ the implied volatility for calls under the rough Heston model with parameters Ξ. Given n call options we define:

$$
r_i(\Xi) := \sigma(\Xi; K_i, T_i) - \sigma^*(K_i, T_i) \qquad i = 1, ..., n
$$

and the residual vector $r(\Xi) = [r_1(\Xi), ..., r_n(\Xi)]^\top$. Then we have to remember that the parameters are subjected to certain constraints: $H \in (0, 1/2)$, $\theta > 0$ and $\rho \in [-1, 1]$. In **Theorem 3.** we used the additional hypothesis $\rho \in (-1/\sqrt{2}, 1/\sqrt{2})$, however, this is not strictly necessary and using the forests approach is possible to obtain the same result with $\rho \in [-1, 1]$. With this notation the calibration of the Heston model is an inverse problem in the nonlinear least square form as:

$$
\min_{\substack{\Xi \\ H\in(0,1/2) \\ \theta>0 \\ \rho\in[-1,1]}} \frac{1}{2}\|r(\Xi)\|^2
$$

since we suppose to have $n \gg 5$ (where 5 is the number of parameters that we have to determine) it is an overdetermined problem. To tackle this kind of problem we will use the Trust Region Reflective algorithm.

6.1 Trust Region Reflective algorithm

6.1.1 Interior Trust Regions approach and reflective transformation

6.1.2 Outline of the algorithm

7 Simulation

References

1. Gatheral, Jaisson, Rosenbaum. Volatility is rough. ArXiv, 2014.

2. Bouchaud, The endogenous dynamics of markets: price impact and feedback loops. ArXiv, 2010.

3. Brunnermeier, Pedersen. Market Liquidity and Funding Liquidity. The Review of Financial Studies, 2009.

4. El Euch, Fukasawa, Rosenbaum. The microstructural foundations of leverage effect and rough volatility. ArXiv, 2016.

5. Jaisson, Rosenbaum. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. ArXiv, 2015.

6. El Euch, Gatheral, Rosenbaum. Roughening Heston. Risk, 2019.

7. El Euch, Gatheral, Rosenbaum. Perfect hedging in rough Heston models. ArXiv, 2017.

8. El Euch, Rosenbaum. The characteristic function of rough Heston models. ArXiv, 2016.

9. Gatheral, Radoicic. Rational approximation of the rough Heston solution. International Journal of Theoretical and Applied Finance, 2019.

10. Alòs, Gathera, Radoicic. Exponentiation of conditional expectations under stochastic volatility. Quantitative Finance, 2020.

11. Branch, Coleman, Li. A subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. SIAM Journal of scientific computing, 1999.