The quadratic Volterra Heston model

(updated 30/06/24). Consider the following stochastic volatility model

$$\begin{aligned} dS_t &= S_t\sqrt{V_t}dW_t \\ Z_t &= g_0(t) + \int_0^t K(t-s)\sqrt{V_s}dW_s \end{aligned}$$

for some $K \in L^2$, where $V_t = aZ_t^2 + c$ and W is a standard Brownian motion. This is a slight variant of the quadratic rough Heston model in [GJR20] (note we have set the parameter b in the original [GJR20] formulation to zero, since the law of V there only depends on $b - Z_0$).

A typical choice for K is the **Gamma kernel** $K(t) := \frac{1}{\Gamma(\alpha)}t^{\alpha-1}e^{-\lambda t}$ with $\alpha = H + \frac{1}{2}$ ($H \in (0, \frac{1}{2}]$) and $g_0(t) = Z_0 + \theta\lambda^{\frac{1}{2}+H}\int_0^t K(t-s)ds = Z_0 + \theta(1 - \frac{\Gamma(\alpha,t\lambda)}{\Gamma(\alpha)}) \to Z_0 + \theta$ as $t \to \infty$ and $g_0(0) = 0$, which is known as the **quadratic rough Heston** model since in this case we expect that V has **rough** sample paths when $H \in (0, \frac{1}{2})$.

We typically want Z_t to be such that $Z_t < 0$, so Z_t decreases (and hence V_t increases) when S_t decreases so implied volatility is negatively skewed. Note we can re-write the defining eq for V just in terms of V and not Z as

$$V_t = a(g_0(t) + \int_0^t K(t-s)\sqrt{V_s}dW_s)^2 + c = a(g_0(t) + \int_0^t K(t-u)\frac{dS_u}{S_u})^2 + c \tag{1}$$

and since V only depends on the history of S itself, we call this a **pure feedback** model. We can also we add an additional term of the form $f(t) + \int_0^t \kappa(t-s)\sqrt{V_s}dB_s$ to the right hand side of (1) (note this additional term is linear not quadratic in $\int_0^t \kappa(t-s)\sqrt{V_s}dB_s$) with $dB_t dW_t = \rho dt$, which simplifies to the (non-quadratic) **rough Heston model** when $a = 0$.

Simulating VIX_T

We can decompose $Z_{t+\tau}$ as

$$Z_{t+\tau} = g_t(\tau) + \int_t^{t+\tau} K(t+\tau-s)\sqrt{V_s}dW_s$$

where $g_t(\tau) := \mathbb{E}_t(Z_{t+\tau}) = g_0(t+\tau) + \int_0^t K(t+\tau-s)\sqrt{V_s}dW_s$. Then we see that

$$\begin{aligned} \mathbb{E}_t(Z_{t+\tau}^2) &= \mathbb{E}_t(Z_{t+\tau})^2 + \mathrm{Var}_t(Z_{t+\tau}) = \mathbb{E}_t(Z_{t+\tau})^2 + \mathbb{E}_t((\int_t^{t+\tau} K(t+\tau-s)\sqrt{V_s}dW_s)^2) \\ &= g_t(\tau)^2 + \int_t^{t+\tau} K(t+\tau-s)^2\mathbb{E}_t(V_s)ds \\ &= g_t(\tau)^2 + \int_t^{t+\tau} K(\tau-(s-t))^2\mathbb{E}_t(V_s)ds \end{aligned} \tag{2}$$

and $\tilde{\xi}_t(\tau) := \mathbb{E}_t(V_{t+\tau})$ is the (shifted) **forward variance curve**. Then we see that

$$\begin{aligned} V_{t+\tau} &= aZ_{t+\tau}^2 + c \\ \tilde{\xi}_t(\tau) = \mathbb{E}_t(V_{t+\tau}) &= \mathbb{E}_t(aZ_{t+\tau}^2 + c) \\ &= ag_t(\tau)^2 + a\int_0^\tau K(\tau-s)^2\tilde{\xi}_t(s)ds + c = f_t(\tau) + a\int_0^\tau K(\tau-s)^2\tilde{\xi}_t(s)ds \end{aligned} \tag{3}$$

where $f_t(\tau) := ag_t(\tau)^2 + c$, and recall that $g_t(\tau)$ is random as it contains a Volterra-type stochastic integral term.

Eq (3) is a linear **Volterra Integral Equation** (VIE) for $\tilde{\xi}_t(.)$ of the form

$$x(\tau) + (k * x)(\tau) = f(\tau) \tag{4}$$

with $x(\tau) = \tilde{\xi}_t(\tau)$, $k(\tau) = -aK(\tau)^2$ and $f(\tau) = f_t(\tau)$ in our case, and the **convolution operator** $*$ is defined by $(f * g)(\tau) := \int_0^\tau f(s)g(\tau-s)ds = \int_0^\tau f(\tau-s)g(s)ds$. The solution to this type of equation is given by

$$x(\tau) = f(\tau) - (r * f)(\tau) \tag{5}$$

where r is the **resolvent** of k, which is the unique function r which satisfies $r + r * k = k$. To see this, we substitute (5) into (4) to get

$$x + k * x = x + k * (f - r * f) = x + (k - k * r) * f = x + r * f = f$$

so we see that (4) is satisfied. Hence the solution to (3) is

$$\tilde{\xi}_t(\tau) \;=\; f_t(\tau) - (\tilde{R} * f_t)(\tau) \;=\; f_t(\tau) - \int_0^\tau \tilde{R}(\tau - s) f_t(s) ds$$

where $\tilde{R}$ is the resolvent of $\tilde{K}(t) := -aK(t)^2$ (see section 6.2 in Rømer[Rom22]). For the choice $K(t) = \frac{1}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}$, we can write

$$\tilde{K}(t) \;=\; -a^2 \frac{1}{\Gamma(\alpha)^2} t^{2\alpha-2} e^{-2\lambda t} \;=\; c^* \frac{1}{\Gamma(\alpha^*)} t^{\alpha^*-1} e^{-\lambda^* t}$$

for some c^*, λ^* and α^*. Then it turns out the resolvent of $\tilde{R}$ can be computed explicitly as

$$c^* e^{-\lambda^* t} t^{\alpha^*-1} E_{\alpha^*,\alpha^*}(-ct^{\alpha^*})$$

where $E_{\alpha,\beta}(t) = \sum_{n=0}^\infty \frac{t^n}{\Gamma(\alpha n + \beta)}$ is known as the **Mittag-Leffler** function. The **Laplace transform** of $\lambda t^{-\alpha-1} E_{\alpha,\alpha}(-\lambda t^\alpha)$ has the simple explicit form $\frac{\lambda}{\lambda + z^\alpha}$.

Finally we compute the VIX index as

$$\begin{aligned}
\mathrm{VIX}_t \;&=\; \frac{1}{\Delta}\mathbb{E}_t(\int_t^{t+\Delta} V_u du) \;=\; \frac{1}{\Delta}(\mathbb{E}_t(\int_0^\Delta V_{t+\tau} d\tau) \;=\; \frac{1}{\Delta}\int_0^\Delta \tilde{\xi}(\tau) d\tau \\
&=\; \frac{1}{\Delta}(\int_0^\Delta (f_t(\tau) d\tau - \int_0^\tau \tilde{R}(\tau - s) f_t(s)(s) ds) d\tau) \\
&=\; \frac{1}{\Delta}(\int_0^\Delta f_t(\tau) d\tau - \int_0^\Delta \int_s^\Delta \tilde{R}(\tau - s) d\tau f_t(s) ds) \\
&=\; \frac{1}{\Delta}(\int_0^\Delta f_t(\tau) d\tau - \int_0^\Delta \int_0^{\Delta - s} \tilde{R}(\tau) d\tau f_t(s) ds) \\
&=\; \frac{1}{\Delta}(\int_0^\Delta f_t(\tau) d\tau - \int_0^\Delta \bar{R}(\Delta - s) f_t(s) ds) \\
&=\; \frac{1}{\Delta}\int_0^\Delta (1 - \bar{R}(\Delta - s)) f_t(s) ds
\end{aligned}$$

where $\bar{R}(t) := \int_0^t R(s) ds$. We would typically compute this integral numerically using **Gaussian quadrature**. Note that VIX_t is a **random variable**, since $f_t(\tau)$ depends on $g_t(\tau)$, and the latter involves a Volterra-type stochastic integral. We can still also use antithetic sampling for VIX_t, i.e. simulate two paths V and $V^{(2)}$ driven by W and $-W$ respectively.

We can approximate V numerically with an **Euler**-type scheme as follows:

$$Z_{j\Delta t} \;=\; g_0(j\Delta t) + \sum_{k=0}^{j-1} K((j-k)\Delta t) \sqrt{V_{k\Delta t}}\, \Delta W_k$$

where ΔW_k is a sequence of i.i.d. $N(0, \Delta t)$ random variables. Since this sum has to be computed for each i, simulating 1 path of Z requires a **double loop**, since we have to repeat this for $j = 1..N$, where N is the number of **time steps** for the Monte Carlo scheme.

We can improve accuracy using

$$\sigma_{j,k}^2 \;=\; \mathbb{E}((\int_{t_k}^{t_{k+1}} K(j\Delta t - s) dW_s)^2) \;=\; \int_{t_k}^{t_{k+1}} K(t-s)^2 ds \;=\; F(t_{k+1}) - F(t_k)$$

(where we have set $t = j\Delta t$) and $F(s) = 4^{-H} \lambda^{-2H} \Gamma(2H, 2(t-s)\lambda)$ (where $\Gamma(a,z) = \int_z^\infty t^{\alpha-1} e^{-t} dt$ here is the **incomplete Gamma function** for which MATLAB and Python have in built functions), and now approximate the Z process as

$$\begin{aligned}
Z_{j\Delta t} \;&=\; g_0(j\Delta t) + \sum_{k=0}^{j-1} \sigma_{j,k} \sqrt{V_{k\Delta t}}\, \tilde{Z}_k \\
V_{j\Delta t} \;&=\; aZ_{j\Delta t}^2 + c
\end{aligned}$$

where the $\tilde{Z}_k$'s are i.i.d standard Normals.

References

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

[Rom22] S.Rømer, "Hybrid multifactor scheme for stochastic Volterra equations with completely monotone kernels", preprint, 2022

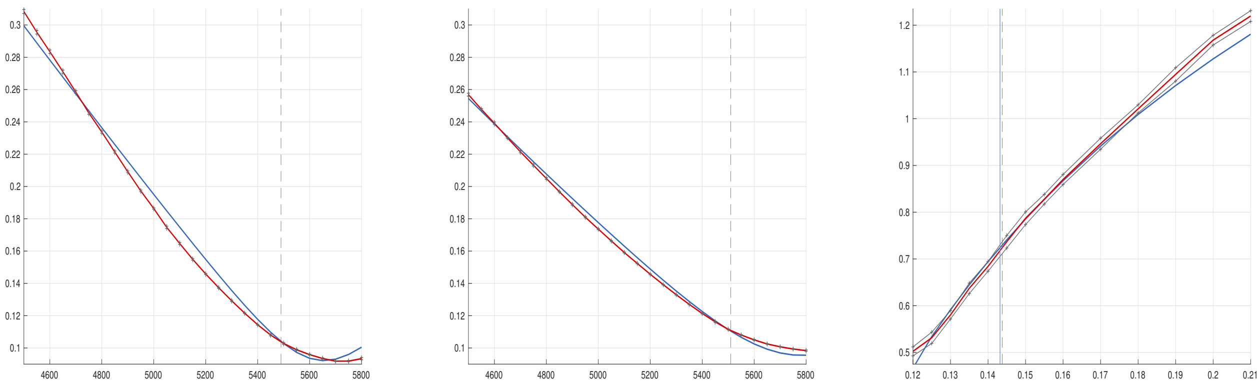

Figure 1: Here we have plotted market mid-implied vol (in red, computed from the average of out-of-the-money bid and ask option prices for which the spreads are tighter) vs calibrated model (blue) implied volatility as a function of the strike K at 1545 EST on Fri 21st June 2024 for SPX smiles with $T = 1$ mnth $= 20/251$ (left), $T = 2$mnths $= 40/251$ (middle), and VIX smile for $T = 1$mnth $= 17/251$ (right) using data from CBOE datashop, calibrated parameters are $H = 0.0624$, $a = 0.321$, $c = 0.00436$, $\lambda = 5.136$, $\theta = -0.0922$ and $Z_0 = -0.0509$ so $\sqrt{V_0} = .0720$ (recall that Z_0 is not a paramater), using 2.5million paths and 2048 time steps with antithetic sampling. The crosses are the bid and ask implied vols, and the dashed vertical lines are the model (blue) and market (grey dashed) VIX forward prices. Note that for the SPX smiles, the qrHeston smile does not lie within the bid-offer spread, but this is the case for the lower strikes of the VIX smile. Calibration does not appear to work so well on Fri 20th Jan 2023 when peculiarly where was little difference between the 1month and 2 month SPX smiles i.e. not much skew flattening, and for this date $H = .132$ which is clearly a lot larger than for the 2024 fits here. We thank John Armstrong for his help in running the Python code for this on a GPU using the CREATE HPC cluster at kcl.

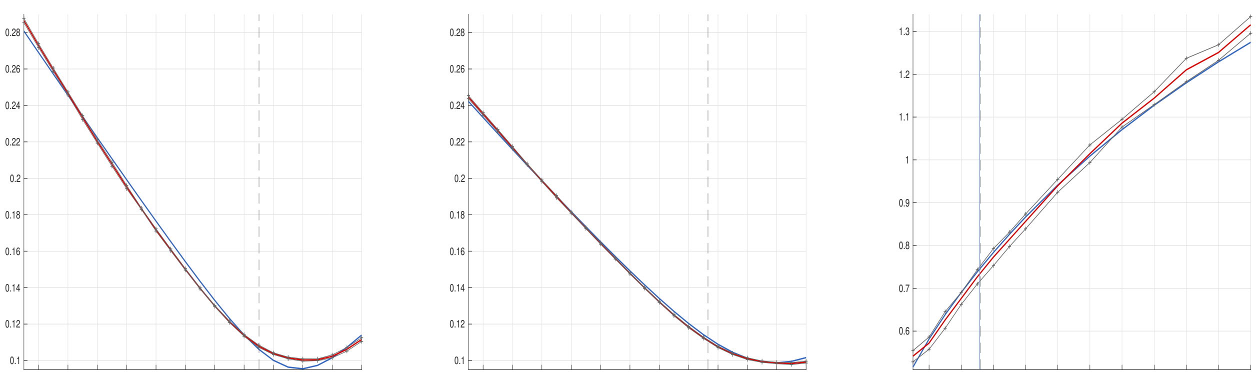

Figure 2: Here we see calibration results for the same maturities on Fri 21st Jan 2024, for which $H = .0671$, $a = 0.337$, $c = 0.00492$, $\lambda = 3.77$, $\theta = -0.0943$ and $Z_0 = -0.0470$.