Deep calibration:

The pointwise approach and old-fashioned SV models revisited

Giacomo Bormetti
Department of Mathematics, Alma Mater University of Bologna
giacomo.bormetti@unibo.it

joint work with

Fabio Baschetti
Scuola Normale Superiore Pisa, Italy
fabio.baschetti@sns.it

Pietro Rossi
Prometeia, Bologna, Italy
pietro.rossi@prometeia.com

Quantitative Finance conference
in honour of Michael Dempster's 85th birthday
12 - 15 April, 2023 - Clare College, Cambridge UK

Motivation

- Appropriate modeling of volatility (historical and implicit) is essential for pricing, hedging, and risk management.
- The goal have been pursued by different communities from different perspectives: Financial Fngineering, Statistical Mechanics, Mathematical Finance, and Econometrics.
- **A Physics-inspired attitude: first identify the empirical regularities and then try to model them.**
- Recent years have seen the introduction of more and more realistic models whose analytical treatment is not always viable and when it is, is numerically challenging:
  - **rough Heston and Volterra processes (solution of fractional equation)**
  - **quadratic rough Heston (Markovian approximations)**
  - **rough Bergomi and path-dependent volatility models (pure Monte Carlo)**
- The rise of machine learning techniques

Outline of the talk

- **I Part**: revisiting and rehabilitating the **pointwise Neural Network approach** to pricing
- **Interlude**: training with SINC
- **II Part**: are old-style stochastic volatility models really over?

The pointwise Neural Network approach revisited

The Volatility Surface

S: underlying (SPX)
K: strike
T: maturity
(European) Call Option Price:
$C(T,k) = \mathbb{E}[(s_T - k)^+], \quad s_T = \log\left(\frac{S_T}{S_0}\right) \; k = \log\left(\frac{K}{S_0}\right)$
$\sigma_{BS}(T,k)$: Black-Scholes implied volatility $C_{mkt}(T,k) = C_{BS}(T,k,\sigma_{BS}(T,k))$

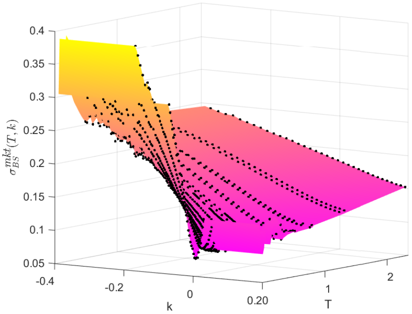


Figure: Market volatility surface as of 2013/08/14

Calibration

Let $\mathcal{M} \equiv \{\mathcal{M}(\theta)\}_{\theta \in \Theta \subset \mathbb{R}^d}$ denote a parametric model of the price evolution. Then, calibration is the numerical procedure identifying the optimal parameters solving

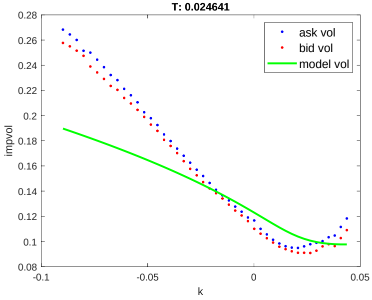
$$\hat{\theta} = \underset{\theta \in \Theta \subset \mathbb{R}^d}{\operatorname{argmin}} \sum_i \sum_j w_{ij} [\sigma_{BS}(T_i, k_j) - \sigma_{BS}^{\mathcal{M}}(T_i, k_j, \theta)]^2$$

where

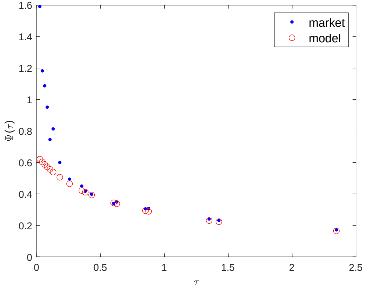
- w_{ij}: set of weights
- $\sigma_{BS}^{\mathcal{M}}(\cdot, \cdot, \theta)$: model implied volatilities for parameter θ (over the market grid in (T, k))
- $\sigma_{BS}(\cdot, \cdot)$: market implied volatilies from the volatility surface

The Heston stochastic volatility model

$$\frac{dS_t}{S_t} = \sqrt{v_t}(\rho dW_t + \sqrt{1-\rho^2}dW_t^{\perp})$$
$$dv_t = \kappa(\bar{v} - v_t)dt + \eta\sqrt{v_t}dW_t$$



(a) Heston fit over the shortest expiration as of 2013/08/14



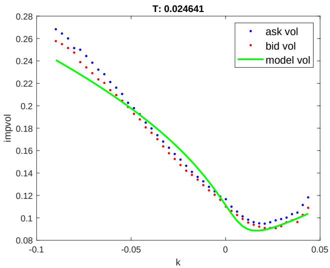
(b) Heston vs term structure of ATM skew as of 2013/08/14

The Rough Heston model

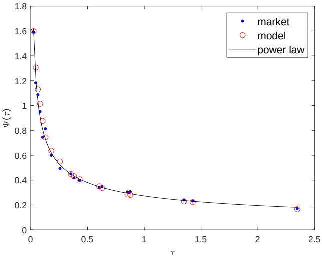
El Euch and Rosenbaum (2018,2019), El Euch, Gatheral, and Rosenbaum *Risk* (2019)

$$dS_t = \sqrt{v_t} S_t (\rho dW_t + \sqrt{1-\rho^2} dW_t^{\perp})$$

$$v_t = \xi_0(t) + \frac{\eta}{\Gamma(H+\frac{1}{2})} \int_0^t \frac{\sqrt{v_s}}{(t-s)^{\frac{1}{2}-H}} dW_s$$



(a) rHeston fit over the shortest expiration as of 2013/08/14



(b) rHeston vs term structure of ATM skew vs power-law fit [$H = 0.0192$] as of 2013/08/14

The forward variance curve, $\xi_0(t)$, is a state-variable, possibly exogenous to calibration.

Two possible routes to calibration

- If the characteristic function (CF) is available in (semi-)closed form:

PROS very efficient machinery to perform computations in Fourier space [FFT-Lewis, SINC from Baschetti et al. *Quant. Finance* (2022)]

CONS the computation of the CF may be slow e.g., rHeston from El Euch and Rosenbaum *Math. Finance* (2019) requires solving a fractional equation. Of course, the rational approximation by Gatheral and Radoičić *IJTAF* (2019) speeds up the procedure.

SOL **one may train a Neural Network to learn the pricing function** to achieve super-fast computation of the implied volatilities and sensitivities

- The model is purely simulative (e.g., rBergomi by Bayer, Priz, Gatheral, *Quant. Finance* (2016) or quadratic rHeston by Gatheral, Jusselin, and Rosenbaum *Risk* (2020), PDV models)

PROS very flexible

CONS Monte Carlo (MC) pricing is computationally too heavy

CONS convergence of the optimization severely affected by MC errors

SOL calibration by feed-forward Neural Networks (Hernandez *Risk* (2017), Horvath, Muguruza, and Tomas *Quant. Finance* (2021))

Calibration by feed-forward Neural Networks

Grid-based methods in the literature (Hernandez (2017), Horvath *et al.* (2021), Rømer *Quant. Finance* (2022))

Train the Neural Network to learn a collection of pixels from the volatility surface over a *fixed* bi-dimensional grid in (T, k)

k/T	T_1	T_2	$\dots$	T_i	$\dots$	T_n
k_1	$\cdot$	$\cdot$	$\dots$	$\cdot$	$\dots$	$\cdot$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k_j	$\cdot$	$\cdot$	$\dots$	$\cdot$	$\dots$	$\cdot$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k_m	$\cdot$	$\cdot$	$\dots$	$\cdot$	$\dots$	$\cdot$

The 2-step approach

- a Neural Network approximates the map from model parameters to (reduced) volatility surfaces (**offline training**)

input	output
θ	$(\sigma_{BS}^{\mathcal{M}}(T_i, k_j, \theta))_{i=1,\dots,n \;\; j=1,\dots,m}$

- solve the optimization problem **online**

Calibration by feed-forward Neural Networks cnt'd

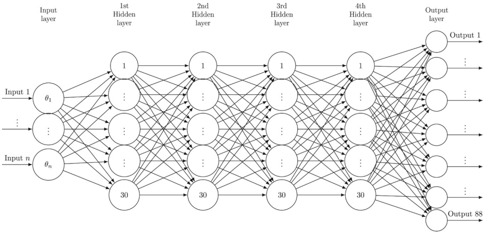


Figure: Neural network architecture in Horvath et al. (2021).

PROS **calibration is extremely fast!** One millisecond to compute one 11×8 grid of implied volatilities, nearly 100 steps to convergence

PROS Rømer's adaptive grid enhances the performances

CONS **need for an interpolation/extrapolation method to compute the implied volatility for arbitrary** (k, T)

CONS **extrapolation, expecially at very low time to maturity, may be subtle**

The pointwise approach

- A feed-forward neural network is trained to learn the pricing function from model parameters and option's parameters to (a single point in) implied volatility

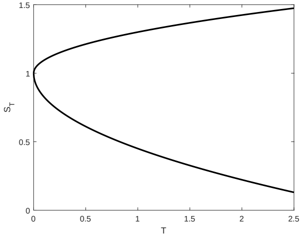
input	output
(T, k, θ)	$\sigma_{BS}^{\mathcal{M}}(T, k, \theta)$

- The approach was pioneered by Bayer and Stemper (2018). **By design, it is interpolation-free.** Unfortunately, the architecture in the original specification was quite demanding: from few hundreds up to four thousands nodes per layer. The number of samples must be large for proper training.
- Horvath et al. (2018) support the grid-based approach since "*by evaluating the objective function on a larger set of (grid) points, injectivity of the mapping can be more easily guaranteed than in the pointwise training*"
- Remark: The mapping from model parameters, strike, and time to maturity is not injective. *Proper training is much more crucial.* A suitable choice of the target will fix injectivity (parameter identification).

Our contribution

We identify three important ingredients:

1. An **adaptive grid**:

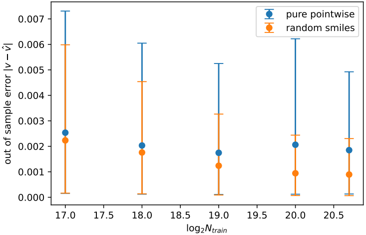


2. The following parametric form for the forward variance curve (vs the piecewise constant specification) performs very well:

$$\xi_0(t) = \beta_0 + \beta_1 \exp\left(-\frac{t}{\tau_1}\right) + \beta_2\left(\frac{t}{\tau_2}\right)\exp\left(-\frac{t}{\tau_2}\right)$$

Our contribution

3. **Pointwise training can be performed more effectively**: One produces **random grid surfaces** in the generation phase **but feed them to the network in a pointwise manner**



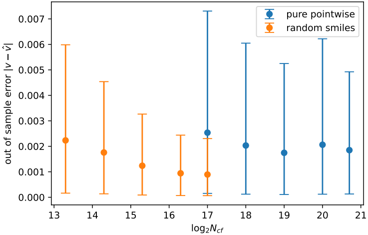


Figure: Out-of-sample errors as a function of the training sample dimension (left) and of the computational burden (right). Blue bars: pointwise training. Orange bars: random smiles training.

4. Network's architecture is as simple as Horvath et al (2018)

March 16, 2021 example

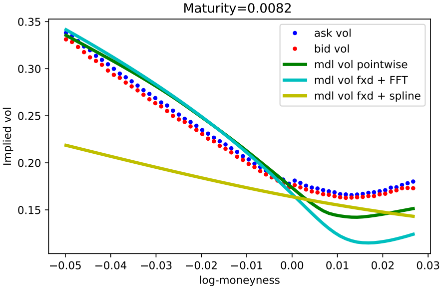


Figure: Rough Heston calibration to the market volatility surface as of March 16, 2021. Green is pointwise (random grid), cyan is fixed grid (interpolation via FFT of the true CF), and yellow is fixed grid plus splines.

July 17, 2014 example

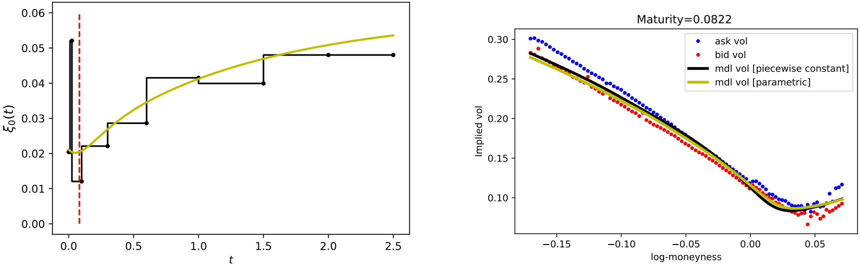


Figure: Left panel: Calibrated piecewise constant (black) and parametric (yellow) forward variance curve as of 2014/07/17. Right panel: fit to the $T = 0.0822$ maturity smile.

Interlude

Random smiles with SINC

The SINC approach

F.Baschetti, G.Bormetti, S.Romagnoli, and P.Rossi, *Quant. Finance* **22**(3), 427 – 446, (2022)

Decompose a European put into Cash or Nothing (CoN) and Asset or Nothing (AoN) options:

$$\mathbb{E}[(K - S_T)^+] = K\mathbb{E}[\mathbb{1}_{s_T<k}] - S_0\mathbb{E}^{s_T}[\mathbb{1}_{s_T<k}]$$

where $s_T = \log \frac{S_T}{S_0}$ and $k = \log \frac{K}{S_0}$.

$$\mathbb{E}[\mathbb{1}_{s_T<k}] = \int f(s_T)\theta(k - s_T)ds_T = \int e^{-i2\pi k\omega}\hat{f}(\omega)\frac{i}{2\pi}\frac{1}{\omega + i\epsilon}d\omega$$

$$\mathbb{E}^{s_T}[\mathbb{1}_{s_T<k}] = \int e^{-i2\pi k\omega}\hat{h}(\omega)\frac{i}{2\pi}\frac{1}{\omega + i\epsilon}d\omega$$

where

$$\hat{h}(\omega) = \int e^{i2\pi\omega s_T}e^{s_T}f(s_T)ds_T = \hat{f}\left(\omega - \frac{i}{2\pi}\right)$$

Shannon Sampling Theorem (SST)

Sufficient condition for a sample rate to guarantee that a discrete sequence of samples resolves all the frequency content and perfectly reproduces the original function.

Assumption The input signal has finite bandwidth in the frequency domain

Inverse problem What if the original function is limited in the direct space?

For any $\eta > 0$, we can find X_c such that

$$\left| 1 - \int_{-X_c}^{X_c} f(s_T) ds_T \right| < \eta$$

and the SST guarantees that the Fourier transform of the truncated PDF $f(s_T)\mathbb{1}_{-X_c < s_T < X_c}$ can be recovered given a countable set of points $\omega_n = \frac{n}{2X_c}$:

$$e^{-i2\pi k\omega} \widehat{f\mathbb{1}_{-X_c<s_T<X_c}}(\omega) = \sum_{n=-\infty}^{+\infty} e^{-i2\pi k\omega_n} \widehat{f\mathbb{1}_{-X_c<s_T<X_c}}(\omega_n) \text{sinc}[2\pi X_c(\omega - \omega_n)]$$

$$\mathbb{E}[\mathbb{1}_{s_T<k}] \approx \mathbb{E}[\mathbb{1}_{s_T<k}\mathbb{1}_{-X_c<s_T<X_c}]$$

$$= \frac{i}{2\pi} \sum_{n=-\infty}^{+\infty} e^{-i2\pi k\omega_n} \widehat{f\mathbb{1}_{-X_c<s_T<X_c}}(\omega_n) \int_{-\infty}^{+\infty} \frac{\text{sinc}[2\pi X_c(\omega - \omega_n)]}{\omega + i\epsilon} d\omega$$

Solving the inner integral in the sinc and truncating the infinite sum

$$\mathbb{E}[\mathbb{1}_{s_T<k}] \approx \frac{i}{2\pi} \sum_{n=-N/2}^{+N/2} e^{-i2\pi k\omega_n} \widehat{f\mathbb{1}_{-X_c<s_T<X_c}}(\omega_n) \left[-i\pi \mathbb{1}_{n=0} + \frac{1-(-1)^n}{n} \mathbb{1}_{n\neq 0} \right]$$

Weights follow a very simple pattern!

The final formula follows replacing $\widehat{f\mathbb{1}_{-X_c<s_T<X_c}}(\omega_n)$ with $\hat{f}(\omega_n)$ and recognizing that even Fourier moments are not relevant:

$$\mathbb{E}[\mathbb{1}_{s_T<k}] \approx \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{N/4} \frac{1}{2n-1} \Big[\sin(2\pi k\omega_{2n-1}) \Re[\hat{f}(\omega_{2n-1})] + - \cos(2\pi k\omega_{2n-1}) \Im[\hat{f}(\omega_{2n-1})] \Big]$$

Out of the $N+1$ terms included in the expansion, only $N/4$ survive!

Proposition (Error analysis)

The error associated to the SINC approach can be written as a sum of three components:

1. the PDF truncation error
2. the approximation to a finite sum
3. true Fourier coefficients for the original PDF in place of the truncated one

$$\begin{aligned}
\epsilon \;=\; & \int f(s_T)\theta(k-s_T)\,ds_T - \int_{-X_c}^{X_c} f(s_T)\theta(k-s_T)\,ds_T \qquad \left(\underset{X_c\to\infty}{\longrightarrow} 0\right) \\
+ \; & \frac{i}{2\pi X_c} \sum_{|n|>N/4} e^{-i2\pi k\omega_{2n-1}} \frac{\widehat{f\mathbb{1}_{-X_c\le s_T\le X_c}}(\omega_{2n-1})}{\omega_{2n-1}} \qquad \left(\underset{N\to\infty}{\longrightarrow} 0\right) \\
+ \; & \frac{i}{2\pi X_c} \sum_{n=-N/4}^{+N/4} e^{-i2\pi k\omega_{2n-1}} \frac{\widehat{f\mathbb{1}_{-X_c\le s_T\le X_c}}(\omega_{2n-1}) - \hat{f}(\omega_{2n-1})}{\omega_{2n-1}} \\
& \left(\le C\frac{X_c}{\pi}\left(f(X_c)+f(-X_c)\right) \underset{X_c\to\infty}{\longrightarrow} 0\right)
\end{aligned}$$

SINC via Fast Fourier Transform

Focusing on a discrete grid of evenly spaced strikes

$$k_m = m\frac{2X_c}{N} \quad -N/2 \leq m < N/2$$

one can write the AoN and CoN prices as

$$\mathbb{E}[e^{a s_T} \mathbb{1}_{s_T < k_m}] \approx$$

$$\frac{i}{2\pi} \sum_{n=0}^{N-1} e^{-i \frac{2\pi}{N} mn} q_n, \qquad q_n = \begin{cases} \frac{\pi}{i} & n = 0 \\ \hat{f}(\omega_n - a\frac{i}{2\pi})\frac{1-(-1)^n}{n} & n \in [1, N/2) \\ 0 & n = N/2 \\ \hat{f}(\omega_{n-N} - a\frac{i}{2\pi})\frac{1-(-1)^{n-N}}{n-N} & n \in (N/2, N-1] \end{cases}$$

where $a = 0$ for the CoN or $a = 1$ for the AoN.

Are old-style stochastic volatility models really over?

Motivation

*An ambitious modeling of volatility requires **first identifying the empirical phenomena** that fundamentally characterize the volatility formation process.*
Slides by L.Parent (2023) available from SSRN 4396307

$$\frac{\mathrm{d}S(t)}{S(t)} = f(t, Y(t))\mathrm{d}W_1(t)$$

$$Y(t) = Y_0 + \int_0^t \mu(t, u, Y(u))\mathrm{d}u + \int_0^t \nu(t, u, Y(u)) \left(\rho \frac{1}{f(u, Y(u))} \frac{\mathrm{d}S(u)}{S(u)} + \sqrt{1 - \rho^2}\mathrm{d}W_u^{\perp} \right)$$

- $\rho \neq \{-1, 0, 1\}$: **partial path-dependency**
- if $f(t, Y(t)) = \sqrt{Y(t)}$ and $\mu(t, u, Y(u))$, $\nu(t, u, Y(u))$ properly specified: **rough Heston model** (El Euch and Rosenbaum (2018)) -> roughness and volatility persistence
- if $f(t, Y(t)) = \sqrt{a + b(Y(t) - c)^2})$ and $\mu(t, u, Y(u))$, $\nu(t, u, Y(u))$ properly specified: **quadratic rough Heston model** (Gatheral *et al.* (2020)) -> Zumbach effect

Volatility is (mostly) path-dependent

Guyon and Lekeufack (2022). Available at SSRN4174589.

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_1(t)$$

$$\sigma_t = \beta_0 + \beta_1 m_1(t) + \beta_2 \sqrt{m_2(t)} + \beta_{1,2}(m_1'(t))^2 \mathbf{1}_{m_1'(t)>0}$$

$$m_1(t) = \int_{-\infty}^{t} K_1(t-u) \frac{\mathrm{d}S(u)}{S(u)}$$

$$m_2(t) = \int_{-\infty}^{t} K_2(t-u) \sigma_u^2 \mathrm{d}u$$

where $m_1'(t)$ is similar to $m_1(t)$ but with a possible different time scale.

- σ_t **purely endogenous** and driven by $m_1(t)$
- lack of probability on the right tail (misspricing of ATM calls):

  $$\mathrm{d}S(t)/S(t) = \mu_t \mathrm{d}t + \mathrm{e}^{X_t} \sigma_t \mathrm{d}W_1(t)$$

  with X_t Ornstein-Uhlenbeck (OU): **mostly** path-dependent
- $\beta_{1,2}$ very significant when pricing

The rough path-dependent volatility model

L. Parent (2023). Available at SSRN4270481.

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_1(t)$$

$$(\sigma_t)^p = \beta_0 + \beta_1^+((m_1(t) - \bar{m}_1)^+)^{a_1} + \beta_1^-((\bar{m}_1 - m_1(t))^+)^{a_2} + \beta_2(m_2(t))^{\frac{p}{a_2}}$$

$$m_1(t) = c\int_{-\infty}^{t} K_1(t-u)\frac{\mathrm{d}S(u)}{S(u)} + \kappa_1 \int_{-\infty}^{t} K_1(t-u)(\theta_1(u) - m_1(u))\mathrm{d}u$$

$$m_2(t) = \int_{-\infty}^{t} K_2(t-u)(\sigma_u)^{a_2}\mathrm{d}u + \kappa_2 \int_{-\infty}^{t} K_2(t-u)(\theta_2(u) - m_2(u))\mathrm{d}u$$

where $a_1, a_2 \in \{1, 2\}$, and θ_j are $\mathcal{F}_t$-adapted processes.

- $m_2(t)$ is a **market activity**, possible MA of past volatility
- if θ_1 and θ_2 are deterministic, purely endogenous dynamics

Volatility is (mostly) path-dependent: A remark

Guyon and Lekeufack (2022). Available at SSRN4174589.

From the empirical analysis, regressing RV on $R_1(t) \doteq \sum_{t_i \leq t} K_1(t - t_i) r_{t_i}$ and $R_2(t) \doteq \sum_{t_i \leq t} K_2(t - t_i) r_{t_i}^2$:

$$RV(t) = \beta_0 + \beta_1 R_1(t) + \beta_2 \sqrt{R_2(t)}$$

- Leverage and/or trend effect: $\beta_1 < 0$
- Volatility clustering: $\beta_2 \in (0, 1)$
- *"Both factors R_1 and R_2 are needed to satisfactorily explain the volatility"* Guyon and Lekeufack (2022)

However, the continuous time (Markov) specification reads:

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu_t \mathrm{d}t + \sigma(R_1, R_2)\mathrm{d}W_1(t)\,, \quad \sigma(R_1, R_2) \doteq \beta_0 + \beta_1 R_1 + \beta_2 \sqrt{R_2}$$

$$\mathrm{d}R_1(t) = \lambda_1(\sigma(R_1, R_2)\mathrm{d}W_1(t) - R_1(t)\mathrm{d}t)$$

$$\mathrm{d}R_2(t) = \lambda_2(\sigma(R_1, R_2)^2\mathrm{d}t - R_2(t)\mathrm{d}t)$$

<- Key: Brownian QV

What if $\beta_1 = 0$? Deterministic evolution of R_2

Not the case for **observation-driven models in discrete-time**, e.g. GARCH

Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets

S.E. Rømer, *Quantitative Finance* **22**(10), 1805 – 1838, (2022)

Shifted mixture two-factor rough Bergomi type model

$$
\begin{aligned}
\frac{\mathrm{d}S(t)}{S(t)} &= \mu_t \mathrm{d}t + \sqrt{V(t)}\mathrm{d}W_1(t) \\
V(t) &= \zeta_0(t)(\mu X_1(t) + (1-\mu)X_2(t)) + c \\
X_i(t) &= \mathcal{E}(\eta_i(\theta_i Y_1(t) + (1-\theta_i)Y_2(t))) \\
Y_i(t) &= \int_0^t K_i(t-u)\mathrm{d}W_{i+1}(u)
\end{aligned}
$$

- A Neural Network learns - via Monte Carlo - the relation from model parameters to a 175-dimensional **adapted** grid of implied volatility
- Excellent fit to SPX and VIX options, but **one looses a bit the nice connection with the empirical phenomena that characterize the volatility formation process**

The EWMA Heston model

L. Parent, *Quantitative Finance* **23**(1), 71 – 93, (2023)

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu_t \mathrm{d}t + \sqrt{V_t}\mathrm{d}W_1(t)$$

$$\mathrm{d}m(t) = -\frac{1}{\tau_m}\left(\frac{\mathrm{d}S(t)}{S(t)} - m(t)\mathrm{d}t\right)$$

$$\mathrm{d}V(t) = -\frac{1}{\tau_V}(V(t) - \nu_t^2)\mathrm{d}t + \xi\nu_t\sqrt{V(t)}\mathrm{d}W_2(t)$$

with $\nu_t = \nu + (\alpha - \beta m(t))^+$.

- $m(t)$ EWMA of past returns: **trend effect**
- $\tau_m << 1$ *mimics* **roughness**
- ν_t consistent with empirical relation between trend and Realized Volatility (RV)

The EWMA Heston model cnt'd

L. Parent, *Quantitative Finance* **23**(1), 71 – 93, (2023)

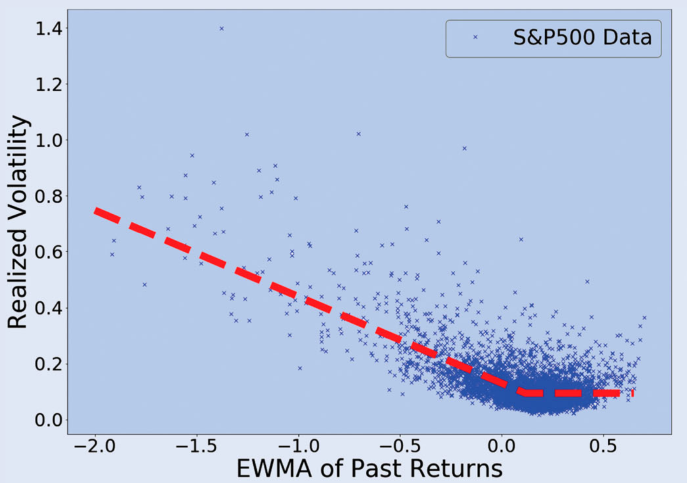


Figure: Taken from Parent, QF (2023). The empirical relationship between the realized volatility of the S&P500 and the EWMA of past returns.

Old-fashioned multi-factor stochastic volatility

Multiple time scales in volatility and leverage correlations

J.Perelló, J.Masoliver, & J-Ph.Bouchaud (2004) *App. Math. Finance*, **11**(1), 27 – 50.

$$\begin{aligned} \mathrm{d}S(t)/S(t) &= \sigma(t)\mathrm{d}W_1(t) \\ \mathrm{d}\sigma(t) &= -\alpha(\sigma(t) - m(t))\mathrm{d}t + k\mathrm{d}W_2(t) \\ \mathrm{d}m(t) &= -\alpha_0(m(t) - m_0)\mathrm{d}t + k_0\mathrm{d}W_3(t) \end{aligned}$$

- Leverage: exponential scaling
- Volatility autocorrelation with multiple time scales

Stochastic volatility with heterogeneous time scales

D. Delpini and G. Bormetti (2015) *Quant. Finance*, **15**(10), 1597 – 1608.

$$\begin{aligned} \mathrm{d}S(t)/S(t) &= (Y(t) + Z(t))\mathrm{d}W_1(t) \\ \mathrm{d}Y(t) &= -\kappa_Y(Y(t) - Y_\infty)\mathrm{d}t + \nu_Y Y(t)\mathrm{d}W_2(t) \\ \mathrm{d}Z(t) &= -\kappa_Z(Z(t) - Z_\infty)\mathrm{d}t + \nu_Z Z(t)\mathrm{d}W_3(t) \end{aligned}$$

- Emergence of power law tails
- Multiple volatility components $\sigma(t) = Y(t) + Z(t)$ mimic long-memory

The EWMA multi-factor exponential OU

$$
\begin{aligned}
\mathrm{d}S(t)/S(t) &= (f(Y(t)) + f(Z(t)))\mathrm{d}W_1(t) \\
\mathrm{d}m_1(t) &= -\frac{1}{\tau_1}(\mathrm{d}S(t)/S(t) - m_1(t)\mathrm{d}t) \\
\mathrm{d}Y(t) &= -\kappa_Y(Y(t) - \log(\nu_1 + (\alpha_1 - \beta_1 m_1(t))^+))\mathrm{d}t + \nu_Y \mathrm{d}W_2(t) \\
\mathrm{d}m_2(t) &= -\frac{1}{\tau_2}(\mathrm{d}S(t)/S(t) - m_2(t)\mathrm{d}t) \\
\mathrm{d}Z(t) &= -\kappa_Z(Z(t) - \log(m_2(t)^2))\mathrm{d}t + \nu_Z \mathrm{d}W_3(t)
\end{aligned}
$$

1. $f(Y(t)) + f(Z(t))$: $f(X) = \mathrm{e}^X$ or $f(X) = X + \sqrt{1 + X^2}$ (same 2^{nd}-order Taylor expansion)
2. τ_1, τ_2: $m_1(t)$ and $m_2(t)$ are EWMAs with different time scales. **Strong Zumbach effect, signed and not-signed.** If $\tau_1 = \tau_2 = +\infty$ exponential OU.
3. *Leverage* from correlation between Brownians and *volatility persistence* from factor structure.

Preliminar Monte Carlo evidence

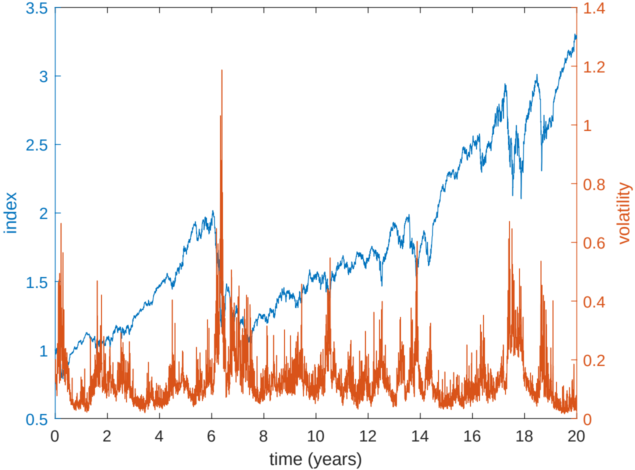


Figure: Returns vs volatility dynamics.

Preliminar Monte Carlo evidence cnt'd

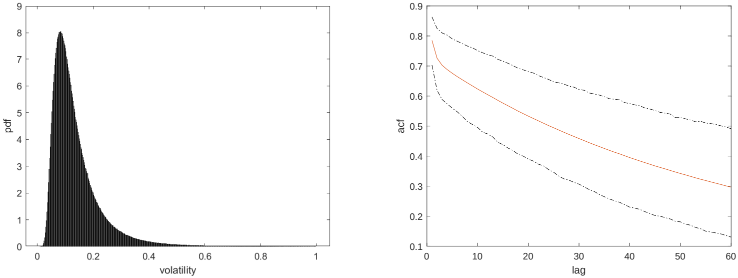


Figure: Left panel: Volatility PDF. Right panel: Volatility serial correlation.

Preliminar Monte Carlo evidence cnt'd

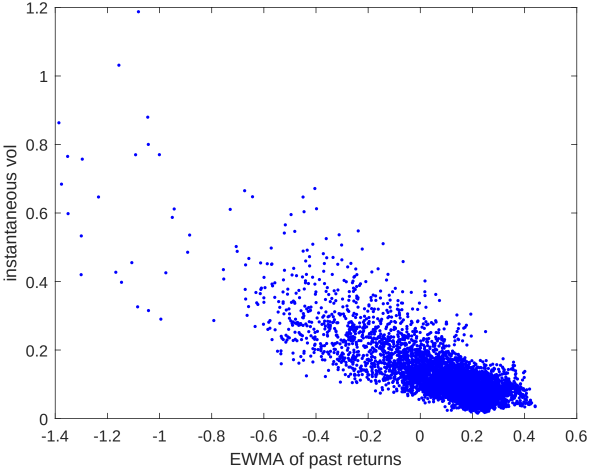
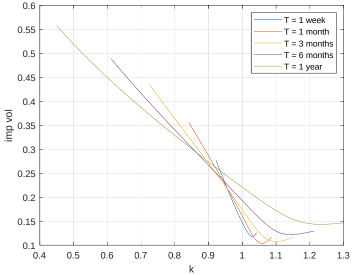
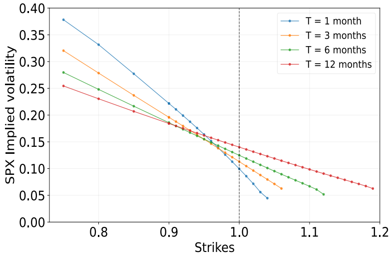


Figure: Spot volatility vs EWMA of past returns.

Preliminar Monte Carlo evidence cnt'd





(A) SPX smiles

Figure: Implied volatility smiles from the model (left panel) and from PDV (right panel, from Guyon and Lekeufack (2022)).

Conclusions and perspectives

- We revisited the pointwise approach to calibration:
  - highly accurate pricing
  - fast computation of the implied volatility and sensitivities
  - **no need for interpolation and extrapolation**
  - **key ingredients**: **adaptive grid and random training with smiles**
- Tested SINC as a smile generator during training
- Proposed a novel (mostly) path-dependent model with Zumbach effect incorporated and multiple volatility time scales (with fast and slow mean reversion)
- What remains to be done: Monte Carlo training to learn the pricing of SPX and VIX options