Fractional volatility models

Jim Gatheral
(joint work with Thibault Jaisson)

(and Christian Bayer, Peter Friz, Andrew Lesniewski, Mathieu Rosenbaum)

Baruch College
The City University of New York

BBQ seminar, Bloomberg New York
Thursday June 26, 2014

Outline of this talk

- Motivation for fractional models
- Empirical volatility statistics
- Fractional Brownian motion (fBm)
- Prior fractional models of volatility
- A data-driven stochastic volatility model: fBergomi
- Fractional Stein Stein and fractional SABR models

Motivation I: Better fitting stochastic volatility models

- Conventional stochastic volatility models generate volatility surfaces that are inconsistent with the observed volatility surface.
  - In stochastic volatility models, the ATM volatility skew is constant for short dates and inversely proportional to T for long dates.
  - Empirically, we find that the term structure of ATM skew is proportional to $1/T^{\alpha}$ for some $0 < \alpha < 1/2$ over a very wide range of expirations.
- The conventional solution is to introduce more volatility factors, as for example in the DMR and Bergomi models.
- One could imagine the power-law decay of ATM skew to be the result of adding (or averaging) many sub-processes, each of which is characteristic of a trading style with a particular time horizon.

Fitting the term structure of ATM skew

- According to (3.21) of [The Volatility Surface], the term structure of ATM skew in a conventional one-factor stochastic volatility model is roughly proportional to

$$\psi(\kappa, \tau) := \frac{1}{\kappa\,\tau}\left\{1 - \frac{1 - e^{-\kappa\,\tau}}{\kappa\,\tau}\right\}.$$

- In Figure 1, we show that this function cannot fit the empirically observed term structure of ATM skew but that adding another such term (as a proxy for adding another factor) generates an excellent fit.

Empirical SPX ATM skew term structure with fits

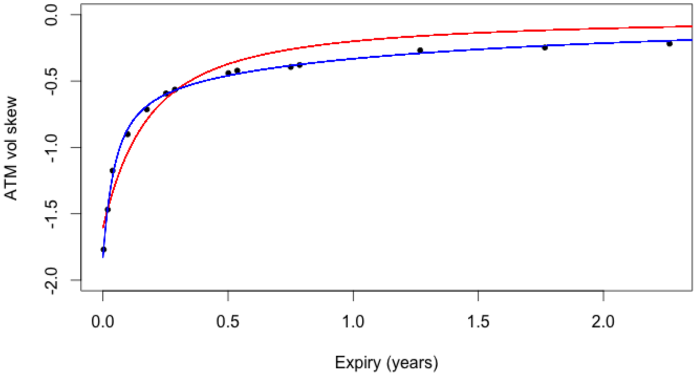

Figure 1 : The points in black are SPX ATM skews as of Sep 15, 2011. The red line is the best fit of $A\,\psi(\kappa,\tau)$. The blue line is the best fit of $A_1\,\psi(\kappa_1,\tau)+A_2\,\psi(\kappa_2,\tau)$.

Bergomi Guyon

- Define the forward variance curve $\xi_t(u) = \mathbb{E}\left[v_u \middle| \mathcal{F}_t \right]$.
- According to [Bergomi and Guyon], in the context of a variance curve model, implied volatility may be expanded as

$$\sigma_{\rm BS}(k,T) = \sigma_0(T) + \sqrt{\frac{w}{T}}\,\frac{1}{2\,w^2}\,C^{x\,\xi}\,k + O(\eta^2) \tag{1}$$

where η is volatility of volatility, $w = \int_0^T \xi_0(s)\,ds$ is total variance to expiration T, and

$$C^{x\,\xi} = \int_0^T dt \int_t^T du\,\frac{\mathbb{E}\left[dx_t\,d\xi_t(u)\right]}{dt}.$$

ATM volatility and the autocorrelation of volatility

- We may write $\xi_t(u) \approx \beta\, \xi_t(t) + \epsilon$ where $\epsilon \perp \xi_t(t)$ and

$$\beta = \frac{\mathrm{cov}(\xi_t(u)\, \xi_t(t))}{\mathrm{var}(\xi_t(t))} = \frac{\mathrm{cov}(v_u, v_t)}{\mathrm{var}(v_t)}$$

which is just the variance autocorrelation $\rho_v(u-t)$.

- Then

$$C^{x\,\xi} \approx \mathbb{E}\left[\frac{\mathbb{E}\left[dx_t\, d\xi_t(t)\right]}{dt}\right] \int_0^T dt \int_t^T du\, \rho_v(u-t).$$

- Thus, the ATM volatility skew

$$\psi(T) := \partial_k \sigma_{\mathrm{BS}}(k,T)|_{k=0} \sim \frac{1}{T^2} \int_0^T dt \int_t^T du\, \rho_v(u-t)$$

which relates the term structure of ATM skew to the variance autocorrelation function.

The Bergomi model

- The n-factor Bergomi variance curve model reads:

$$\xi_t(u) = \xi_0(u) \exp\left\{ \sum_{i=1}^{n} \eta_i \int_0^t e^{-\kappa_i (t-s)} \, dW_s^{(i)} + \text{ drift } \right\}. \tag{2}$$

- To achieve a decent fit to the observed volatility surface, and to control the forward smile, we need at least two factors.
  - In the two-factor case, there are 8 parameters.
- When calibrating, we find that the two-factor Bergomi model is already over-parameterized. Any combination of parameters that gives a roughly $1/\sqrt{T}$ ATM skew fits well enough.
  - Moreover, the calibrated correlations between the Brownian increments $dW_s^{(i)}$ tend to be high.

Tinkering with the Bergomi model

- The term structure of ATM skew is related to the term structure of the autocorrelation function.
- The autocorrelation function of volatility is driven by the exponential kernel in the exponent in (2).
- It's tempting to replace the exponential kernels in (2) with one or more power-law kernels.
- In the single factor case, this would give a model of the form

$$\xi_t(u) = \xi_0(u) \exp\left\{\eta \int_0^t \frac{dW_s}{(t-s)^\gamma} + \text{ drift }\right\}$$

or more generally

$$\xi_t(u) = \xi_0(u) \exp\left\{\eta \int_0^t K(t-s)\, dW_s + \text{ drift }\right\}$$

where the kernel $K(\tau)$ has a power-law singularity as $\tau \to 0$.

Conversely

- Suppose the true model were something like

$$\xi_t(u) = \xi_0(u) \exp\left\{ \eta \int_0^t K(t-s)\, dW_s + \text{ drift } \right\}$$

- Then, using a discrete Laplace transform, we could approximate the kernel as

$$(t-s)^{-\gamma} \approx \sum_{i=1}^{n} \alpha_i \, e^{-\kappa_i\,(t-s)}$$

for some coefficients α_i.

  - Then we would have the Bergomi model back (but with all Brownians perfectly correlated).

Power-laws from averaging: A toy example

The following example, adapted from [Comte and Renault], illustrates how power-law behavior can emerge from the averaging of short memory processes.

- Consider the following OU process ($X_t = \log \sigma_t$ say) indexed by κ:

$$X_t(\kappa) = \int_0^t e^{-\kappa\,(t-s)}\, dW_s.$$

  Then $X_t \sim N(0, \Sigma(\kappa)^2)$ with $\Sigma(\kappa)^2 = \int_0^t e^{-2\,\kappa\,(t-s)}\, ds$.

- Consider a multiplicity of such processes with gamma-distributed κ. Explicitly,

$$p_\Gamma(\kappa) = \frac{\kappa^{\alpha-1}\, e^{-\kappa/\theta}}{\theta^\alpha\, \Gamma(\alpha)}$$

  for some $\alpha > 0$ and $\theta > 0$.

- Then, the average $\bar{X} \sim N(0, \bar{\Sigma}^2)$ with

$$\bar{\Sigma}^2 = \int_0^\infty p_\Gamma(\kappa) \int_0^t e^{-2\,\kappa\,(t-s)}\, d\kappa\, ds = \int_0^t \frac{1}{[1+2\,\theta\,(t-s)]^\alpha}\, ds$$

and

$$\bar{X}_t = \int_0^t \frac{dW_s}{[1+\theta\,(t-s)]^{\alpha/2}}.$$

- Thus, averaging short memory volatility processes (with exponential kernels) over different timescales can generate a volatility process with a power-law kernel

Motivation II: Power-law scaling of the volatility process

- A separate but (presumably) related reason for considering fractional volatility models is that the time series of realized volatility exhibits power-law scaling.
- The Oxford-Man Institute of Quantitative Finance makes historical realized variance (RV) estimates freely available at `http://realized.oxford-man.ox.ac.uk`. These estimates are updated daily.
- Using daily RV estimates as proxies for instantaneous variance, we may investigate the time series properties of v_t empirically.

SPX realized variance from 2000 to 2014

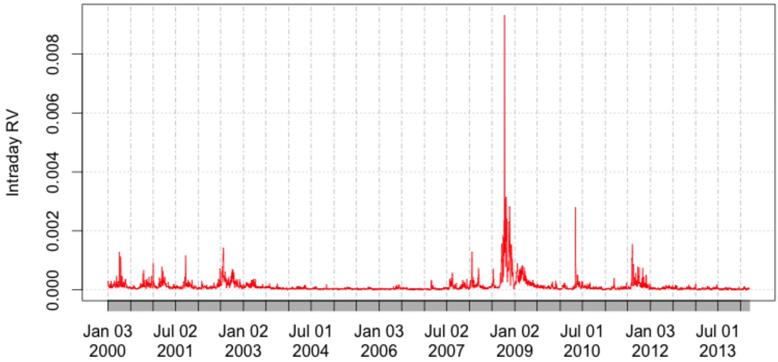

Figure 2 : KRV estimates of SPX realized variance from 2000 to 2014.

The variance of SPX RV

- Let v_t^R denote the realized variance of SPX on day t.
- We consider two measures of variance of RV over the time interval Δ[1]:

1. $$V_1(\Delta) := \langle (v_{t+\Delta}^R - v_t^R)^2 \rangle$$

2. $$V_2(\Delta) := \langle (\log(v_{t+\Delta}^R) - \log(v_t^R))^2 \rangle$$

- We find that

$$V_2(\Delta) = A\,\Delta^{2\,H} \text{ with } H \approx 0.14 \text{ and } A \approx 0.38.$$

[1] $\langle \cdot \rangle$ denotes a sample average.

Stationarity of $\log(v_t^R)$

- Suppose $\log(v_t^R)$ is mean-square stationary for large t (as we certainly believe).
- Then,

$$V_2(\Delta) = \langle (\log(v_{t+\Delta}^R) - \log(v_t^R))^2 \rangle \leq 4\, M_2$$

  where $M_2 = \mathrm{var}(\log v^R) < \infty$.
- If so, we must have $V_2(\Delta) \sim$ const. as $\Delta \to \infty$. Power-law scaling of $V_2(\Delta)$ can hold only up to some long timescale.
  - A hand-waving estimate of this timescale using $A\,\tilde{\Delta}^{2\,H} \approx 4\, M_2$ gives $\tilde{\Delta} \approx 24$ years.

Variance of RV differences vs Δ

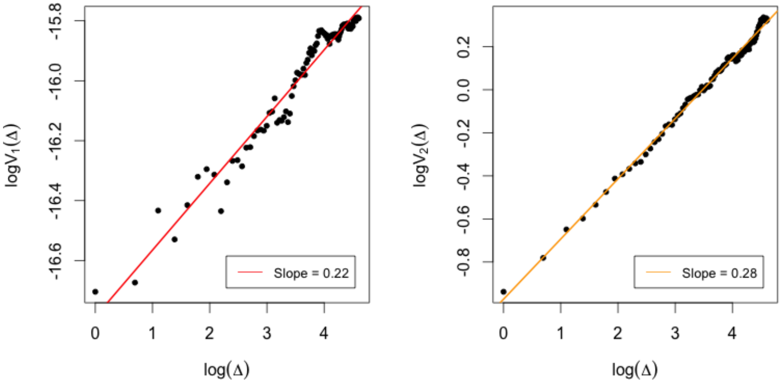

Figure 3 : Log-log plots of $V_1(\Delta)$ and $V_2(\Delta)$ respectively. $V_2(\Delta)$ wins!

Distributions of $(\log v^R_{t+\Delta} - \log v^R_t)$ for various lags Δ

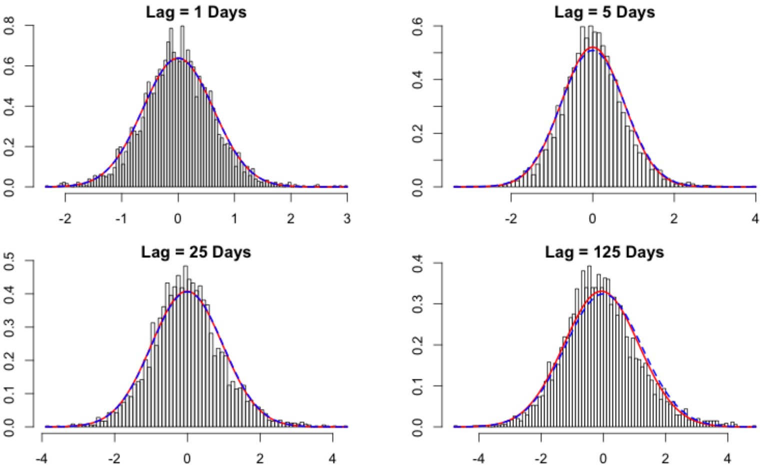

Figure 4 : Histograms of $(\log v^R_{t+\Delta} - \log v^R_t)$ for various lags Δ; normal fit in red; $\Delta = 1$ normal fit scaled by $\Delta^{0.14}$ in blue.

Q-Q plots of $(\log v^R_{t+\Delta} - \log v^R_t)$ for various lags Δ

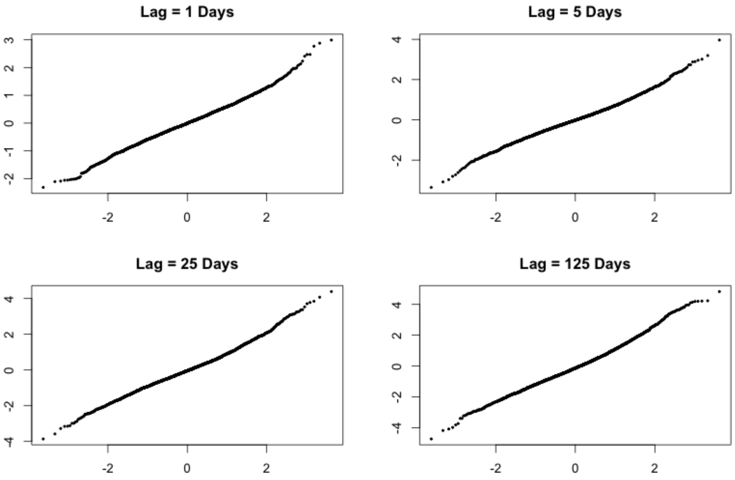

Figure 5 : Q-Q plots of $(\log v^R_{t+\Delta} - \log v^R_t)$ for various lags Δ.

Slopes for all indices

Repeating this analysis for all 21 indices in the Oxford-Man dataset yields:

Index	Slope $= 2\,H$
SPX2.rk	0.28
FTSE2.rk	0.29
N2252.rk	0.22
GDAXI2.rk	0.32
RUT2.rk	0.25
AORD2.rk	0.17
DJI2.rk	0.28
IXIC2.rk	0.26
FCHI2.rk	0.28
HSI2.rk	0.21
KS11.rk	0.26
AEX.rk	0.31
SSMI.rk	0.38
IBEX2.rk	0.27
NSEI.rk	0.23
MXX.rk	0.19
BVSP.rk	0.21
GSPTSE.rk	0.23
STOXX50E.rk	0.27
FTSTI.rk	0.26
FTSEMIB.rk	0.27

Correlogram and test of scaling

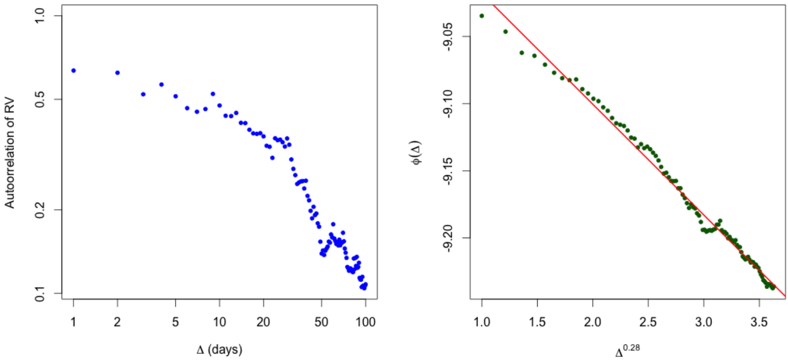

Figure 6 : The LH plot is a conventional correlogram of RV; the RH plot is of $\phi(\Delta) := \langle \log\left(\mathrm{cov}(\sigma_{t+\Delta}, \sigma_t) + \langle \sigma_t \rangle^2\right) \rangle$ vs Δ^{2H} with $H = 0.14$. The RH plot again supports the scaling relationship $V_2(\Delta) \propto \Delta^{2H}$.

A natural model of realized volatility

- Distributions of differences in the log of realized variance are close to Gaussian.
  - This motivates us to model v_t (and so also $\sigma_t = \sqrt{v_t}$) as a lognormal random variable.
- Moreover, the scaling property of variance of RV differences suggests the model:

$$\log v_{t+\Delta} - \log v_t = 2\,\eta\,\left(W^H_{t+\Delta} - W^H_t\right) \tag{3}$$

where W^H is fractional Brownian motion.

Heuristic derivation of autocorrelation function

We assume that $\sigma_t = \bar{\sigma}_t \, e^{\eta \, W^H_t}$. Then

$$
\begin{aligned}
& \operatorname{cov}\left[\sigma_t, \sigma_{t+\Delta}\right] \\
= \quad & \bar{\sigma}_t \, \bar{\sigma}_{t+\Delta} \left[\exp\left\{ \frac{\eta^2}{2} \left(t^{2H} + (t+\Delta)^{2H} - \Delta^{2H} \right) \right\} - 1 \right] \\
\sim \quad & \bar{\sigma}_t \, \bar{\sigma}_{t+\Delta} \, \exp\left\{ \frac{\eta^2}{2} \left(t^{2H} + (t+\Delta)^{2H} - \Delta^{2H} \right) \right\} \text{ as } t \to \infty.
\end{aligned}
$$

Similarly,

$$
\operatorname{var}\left[\sigma_t\right] \sim \bar{\sigma}_t^2 \, \exp\left\{ \eta^2 \, t^{2H} \right\}.
$$

Thus

$$
\rho(\Delta) = \frac{\operatorname{cov}\left[\sigma_t, \sigma_{t+\Delta}\right]}{\sqrt{\operatorname{var}\left[\sigma_t\right] \, \operatorname{var}\left[\sigma_{t+\Delta}\right]}} \sim \exp\left\{ -\frac{\eta^2}{2} \, \Delta^{2H} \right\}.
$$

Model vs empirical autocorrelation functions

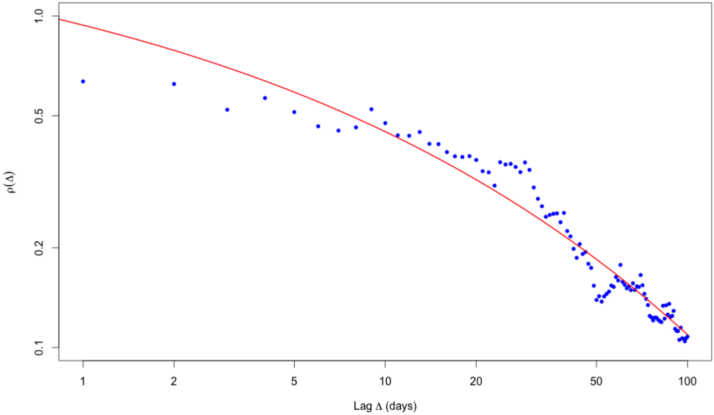

Figure 7 : Here we superimpose the predicted functional form of $\rho(\Delta)$ (in red) on the empirical curve (in blue).

Long memory

- It is a widely-accepted stylized fact that the volatility time series exhibits long memory.
  - For example [Andersen et al.] estimate the degree d of fractional integration from daily realized variance data for the 30 DJIA stocks.
    - Using the DPH estimator, they find d around 0.35 which implies that the ACF $\rho(\tau) \sim \tau^{2\,d-1} = \tau^{-0.3}$ as $\tau \to \infty$.
  - Using the same DPH estimator on the Oxford-Man RV data we find $d = 0.48$. But our model (3) is different from that of [Andersen et al.]. In our case, $\rho(\tau) \sim \tau^{2\,H-1} = \tau^{-0.72}$ as $\tau \to \infty$.
- We see clearly from Figures 4 and 6 that the realized volatility series is "medium memory" with $H \approx 0.14 < 1/2$.

Fractional Brownian motion (fBm)

- *Fractional Brownian motion* (fBm) $\{W_t^H; t \in \mathbb{R}\}$ is the unique Gaussian process with mean zero and autocovariance function

$$\mathbb{E}\left[W_t^H W_s^H\right] = \frac{1}{2}\left\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right\}$$

  where $H \in (0,1)$ is called the *Hurst index* or parameter.
- In particular, when $H = 1/2$, fBm is just Brownian motion.
  - If $H > 1/2$, increments are positively correlated.
  - If $H < 1/2$, increments are negatively correlated.

Representations of fBm

There are infinitely many possible representations of fBm in terms of Brownian motion. For example, with $\gamma = \frac{1}{2} - H$,

Mandelbrot-Van Ness

$$W_t^H = C_H \left\{ \int_{-\infty}^{t} \frac{dW_s}{(t-s)^\gamma} - \int_{-\infty}^{0} \frac{dW_s}{(-s)^\gamma} \right\}.$$

where the choice

$$C_H = \sqrt{\frac{2\,H\,\Gamma(3/2-H)}{\Gamma(H+1/2)\,\Gamma(2-2\,H)}}$$

ensures that

$$\mathbb{E}\left[W_t^H\,W_s^H\right] = \frac{1}{2}\left\{t^{2H} + s^{2H} - |t-s|^{2H}\right\}.$$

Another representation of fBm

Define

$$K_H(t,s) = C_H\, F\left(\gamma, -\gamma, 1-\gamma, 1-\frac{t}{s}\right) \frac{1}{(t-s)^{\gamma}}.$$

where $F(\cdot)$ is Gauss's hypergeometric function. Then, fBm can also be represented as:

Molchan-Golosov

$$W_t^H = \int_0^t K_H(t,s)\, dW_s.$$

- The Mandelbrot-Van Ness representation uses the entire history of the Brownian motion $\{W_s; s \leq t\}$.
- The Molchan-Golosov representation uses only the history of the Brownian motion from time 0.

Why "fractional"?

Denote the differentiation operator $\frac{d}{dt}$ by D. Then

$$D^{-1} f(t) = \int_0^t f(s)\, ds.$$

The Cauchy formula for repeated integration gives for any integer $n > 0$,

$$D^{-n} f(t) = \int_0^t \frac{1}{n!} (t-s)^{n-1} f(s)\, ds.$$

The generalization of this formula to real ν gives the definition of the fractional integral:

$$D^{-\nu} f(t) = \frac{1}{\Gamma(\nu)} \int_0^t (t-s)^{\nu-1} f(s)\, ds.$$

Note in particular that $D^0 f(t) = f(t)$.

Comte and Renault: FSV model

- [Comte and Renault] were perhaps the first to model volatility using fractional Brownian motion.
- In their fractional stochastic volatility (FSV) model,

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \sigma_t \, dZ_t \\
d \log \sigma_t &= -\kappa \left(\log \sigma_t - \theta \right) dt + \gamma \, d\hat{W}_t^H
\end{aligned}
\tag{4}
$$

with

$$
\hat{W}_t^H = \int_0^t \frac{(t-s)^{H-1/2}}{\Gamma(H+1/2)} \, dW_s, \quad 1/2 \leq H < 1
$$

and $\mathbb{E}\left[dW_t \, dZ_t \right] = \rho \, dt$.

- The FSV model is a generalization of the Hull-White stochastic volatility model.

Integral formulation

Solving (4) formally gives

$$\sigma_t = \exp\left\{\theta + e^{-\kappa\,t}\,(\log\sigma_0 - \theta) + \gamma \int_0^t e^{-\kappa\,(t-s)}\,d\hat{W}_s^H\right\}. \quad (5)$$

- $H > 1/2$ to ensure long-memory.
- Stationarity is achieved with the exponential kernel $e^{-\kappa\,(t-s)}$ at the cost of introducing an explicit timescale κ^{-1}.

FSV covariance

Define $y_t = \log \sigma_t$. We have

$$\operatorname{cov}(y_t, y_{t+\Delta}) \propto \int_{-\infty}^{0} e^{\kappa\, s}\, ds \int_{-\infty}^{\Delta} e^{\kappa\,(s'-\Delta)}\, ds'\, |s-s'|^{2\,H-2}.$$

Then $\mathbb{E}\left[(y_{t+\Delta} - y_t)^2\right] = 2\operatorname{var}[y_t] - 2\operatorname{cov}(y_t, y_{t+\Delta})$ where

$$\begin{aligned}&\operatorname{cov}(y_t, y_{t+\Delta})\\ &\propto \frac{e^{-k\Delta}}{2k^{2H}} \int_0^{k\Delta} \frac{e^u\, du}{u^{2-2H}} + \frac{e^{-k\Delta}}{2k^{2H}} \Gamma(2H-1) + \frac{e^{k\Delta}}{2k^{2H}} \int_{k\Delta}^{+\infty} \frac{e^{-u}\, du}{u^{2-2H}}.\end{aligned}$$

Vilela Mendes

- An empirical study of the scaling of volatility estimates by [Vilela Mendes and Oliveira] motivates their *data-reconstructed* model:

$$\begin{aligned} \frac{dS_t}{S_t} &= \sigma_t \, dZ_t \\ \log(\sigma_t) &= \beta + \frac{k}{\delta} \left\{ W_t^H - W_{t-\delta}^H \right\} \end{aligned} \tag{6}$$

or equivalently,

$$\sigma_t = \exp\left\{ \beta + \frac{k}{\delta} \int_0^t \mathbb{1}_{s>t-\delta} \, dW_s^H \right\}.$$

- We note that this model looks very similar to the Comte-Renault model (5).
  - The indicator function kernel acts like the exponential kernel in FSV to force stationarity for long times.

Heston model

- The perennially popular and useful Heston model is given by:

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \sqrt{v_t}\, dZ_t \\
dv_t &= -\kappa\,(v_t - \bar{v})\, dt + \eta\,\sqrt{v_t}\, dW_t
\end{aligned}
\tag{7}
$$

  with $\mathbb{E}\left[dW_t\, dZ_t\right] = \rho\, dt$.

- The variance process v_t is a stationary process with:

$$
\begin{aligned}
\mathbb{E}\left[v(t)\right] &= \bar{v} \\
\mathrm{var}[v_t] &= \frac{\eta^2\, \bar{v}}{2\,\kappa} \\
\mathrm{cov}[v_{t+h}\, v_t] &= \frac{\eta^2\, \bar{v}}{2\,\kappa}\, e^{-\kappa\,|h|}.
\end{aligned}
$$

- The variance autocorrelation function is a decaying exponential and so the variance process is short memory.

Comte, Coutin and Renault: Affine model

In the affine fractional stochastic volatility (AFSV) model of [Comte, Coutin and Renault], the Heston model is extended by writing

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{v_t}\, dZ_t \\ v_t &= \theta + \int_{-\infty}^{t} \frac{(t-s)^{\alpha-1}}{\Gamma(\alpha)}\, y_s\, ds, \quad 0 < \alpha < 1/2 \end{aligned} \tag{8}$$

where y_t solves the CIR SDE

$$dy_t = -\kappa\,(y_t - \bar{v})\, dt + \eta\, \sqrt{y_t}\, dW_t.$$

The constant θ is introduced to break the otherwise tight connection between the mean and the variance of v_t.

Autocovariance computations

- From the definition (8), v_t is a "fractionally integrated" CIR process.
- As in the Heston model, v_t is a stationary process with:

$$
\begin{aligned}
\operatorname{cov}[v_{t+h}\, v_t] &= \frac{\eta^2\, \bar{v}}{2\,\kappa}\, \frac{\Gamma(1-2\,\alpha)}{\Gamma(1-\alpha)\,\Gamma(\alpha)} \int_{-\infty}^{\infty} e^{-\kappa\,|u|}\, |u+h|^{2\,\alpha-1}\, du \\
&\sim \frac{1}{\Gamma(2\,\alpha)}\, (\kappa\, h)^{2\,\alpha-1} \text{ as } h \to \infty
\end{aligned}
$$

– asymptotic power-law decay of the autocorrelation function.

Inherited Markovianity in the AFSV model

- One might expect that expected realized variance (integrated variance) would depend on the whole history $\{v_s;\, s < t\}$ of past instantaneous variance.
- However, [Comte, Coutin and Renault] show that the unexpected component

$$\tilde{w}_t(T) := \int_t^T v_s\, ds - \mathbb{E}\left[\int_t^T v_s\, ds \middle| \mathcal{F}_t\right]$$

of the integrated variance depends on $\mathcal{F}_t$ only through the current state v_t.
- Moreover, $\mathbb{E}\left[\int_t^T v_s\, ds \middle| \mathcal{F}_t\right]$ is nothing other than the variance swap curve as of time t.

Long memory vs realized variance (RV) data

- Notwithstanding that long memory of volatility is widely accepted as a stylized fact, RV data does not have this property.
- In Figure 8 we demonstrate graphically that existing long memory volatility models with $H > 1/2$ are not compatible with the RV data.
    - In the FSV model for example, the autocorrelation function $\rho(\tau) \propto \tau^{2H-2}$. Then, for long memory, we must have $1/2 < H < 1$.
        - For $\Delta \gg 1/\kappa$, stationarity kicks in and $V_2(\Delta)$ tends to a constant as $\Delta \to \infty$.
        - For $\Delta \ll 1/\kappa$, the exponential decay in (5) is not significant and $V_2(\Delta) \propto \Delta^{2H}$.

Incompatibility of long-memory models with RV time series

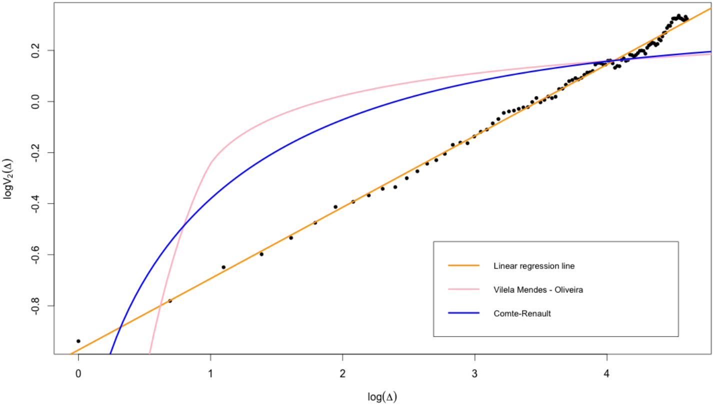

Figure 8 : Long memory models such as Comte-Renault (CR) and Vilela Mendes-Oliveira (VMO) are not compatible with the RV data. The blue line is CR with $k = 0.5$ and $H = 0.53$; the pink line is VMO with $\delta = 1$ day and $H = 0.8$.

The fractional story

Suppose $\log \sigma_t =: y_t$ is given by a Gaussian Volterra process (in the terminology of [Decreusefond]) so that

$$y_t = \eta \int_{-\infty}^{t} K(t,s)\, dW_s \text{ with } K(t,s) \sim \frac{1}{(t-s)^\gamma} \text{ as } s \to t,$$

Then

$$\begin{aligned} y_u - y_t &= \eta \int_t^u K(u,s)\, dW_s + \eta \int_{-\infty}^{t} \left[K(u,s) - K(t,s)\right] dW_s \\ &=: \eta \left(M_{t,u} + Z_{t,u}\right). \end{aligned}$$

Note that $\mathbb{E}\left[M_{t,u} \middle| \mathcal{F}_t \right] = 0$ and $Z_{t,u}$ is $\mathcal{F}_t$-measurable.

- The point is that

$$\mathbb{E}\left[y_u \middle| \mathcal{F}_t \right] - y_t = Z_{t,u}$$

  and $\mathbb{E}\left[y_u \middle| \mathcal{F}_t \right]$ is in principle computable from the volatility surface.

A data-driven stochastic volatility model

- Ignoring (at least for the time being) the difference between historical and pricing measures, we are led naturally from the data to the following model:

$$\log v_u - \log v_t = 2\,\eta\,\left(M_{t,u} + Z_{t,u}\right) \tag{9}$$

- Integrating (9) then gives

$$\begin{aligned} v_u &= v_t \exp\left\{2\,\eta\,M_{t,u} + 2\,\eta\,Z_{t,u}\right\} \\ &= \mathbb{E}\left[v_u \middle| \mathcal{F}_t\right] \exp\left\{2\,\eta \int_t^u K(u,s)\,dW_s - 2\,\eta^2 \int_t^u K(u,s)^2\,ds\right\} \end{aligned} \tag{10}$$

- Of course this could be extended to n factors:

$$v_u = \mathbb{E}\left[v_u \middle| \mathcal{F}_t\right] \exp\left\{2 \int_t^u \sum_{i=1}^n \eta_i\,K^{(i)}(u,s)\,dW_s^{(i)} + \text{ drift}\right\}.$$

  - We could call this a *fractional Bergomi* or *fBergomi* model.

Features of the fractional Bergomi model

- The forward variance curve

$$\xi_t(u) = \mathbb{E}\left[v_u \middle| \mathcal{F}_t \right] = v_t \, \exp\left\{ Z_{t,u} + 2\,\eta^2 \int_t^u K(u,s)^2\, ds \right\}.$$

  depends on the historical path $\{W_s; s < t\}$ of the Brownian motion since inception ($t = -\infty$ say).
- The fractional Bergomi model is non-Markovian:

$$\mathbb{E}\left[v_u \middle| \mathcal{F}_t \right] \neq \mathbb{E}[v_u | v_t].$$

- However, given the (infinite) state vector $\xi_t(u)$, which can in principle be computed from option prices, the dynamics of the model are well-determined.
- In practice, there is a bid-offer spread and we don't have option prices for all strikes and expirations.
  - There is inherent model risk!

Re-interpretation of the conventional Bergomi model

- A conventional n-factor Bergomi model is not self-consistent for an arbitrary choice of the initial forward variance curve $\xi_t(u)$.
    - $\xi_t(u) = \mathbb{E}\left[v_u \middle| \mathcal{F}_t \right]$ should be consistent with the assumed dynamics.
- Viewed from the perspective of the fractional Bergomi model however:
    - The initial curve $\xi_t(u)$ reflects the history $\{W_s; s < t\}$ of the driving Brownian motion up to time t.
    - The exponential kernels in the exponent of (2) approximate more realistic power-law kernels.
- The conventional two-factor Bergomi model is then justified in practice as a tractable Markovian engineering approximation to a more realistic fractional Bergomi model.

From fBergomi to fSABR

- Variance is lognormal in the fBergomi model and thus volatility is also lognormal.
- Then rewrite (10) as

$$\sigma_u = \hat{\sigma}_t(u)\,\mathcal{E}\left(\eta \int_t^u K(u,s)\,dW_s\right) \tag{11}$$

where $\mathcal{E}(X)$ is the stochastic exponential of X and the forward volatility is given by

$$\begin{aligned} \hat{\sigma}_t(u) &= \mathbb{E}\left[\sigma_u \middle| \mathcal{F}_t\right] \\ &= \mathbb{E}\left[\sqrt{v_u} \middle| \mathcal{F}_t\right] \\ &= \sqrt{\xi_t(u)}\,\exp\left\{-\frac{\eta^2}{2}\int_t^u K(u,s)^2\,ds\right\}. \end{aligned}$$

From fBergomi to fSABR

- The formal solution of the fractional Bergomi model may then be written as

$$
\begin{aligned}
S_T &= S_0 \exp\left\{\int_0^T \sigma_t\, dZ_t - \frac{1}{2}\int_0^T \sigma_t^2\, dt\right\} \\
&= S_0\, \mathcal{E}\left(\int_0^T \hat{\sigma}_0(t)\, \mathcal{E}\left(\eta \int_0^t K(t,s)\, dW_s\right)\, dZ_t\right).
\end{aligned}
$$

- $\hat{\sigma}_0(t)$ is typically a slowly-varying function of t so we may write

$$
S_T \approx S_0\, \mathcal{E}\left(\bar{\sigma}(T) \int_0^T \mathcal{E}\left(\eta \int_0^t K(t,s)\, dW_s\right)\, dZ_t\right) \tag{12}
$$

with $\bar{\sigma}(T)^2 = \frac{1}{T}\int_0^T \hat{\sigma}_0(t)^2\, dt$.

The fractional SABR (fSABR) model

Setting $K(t,s) = K_H(t,s)$ (the Molchan-Golosov kernel), we identify (12) as the solution of the following *fSABR* model:

$$
\begin{aligned}
\frac{dS_t}{S_t} &= \sigma_t \, dZ_t \\
\sigma_t &= \bar{\sigma}(T) \, \mathcal{E}\left(\alpha \, W_t^H\right)
\end{aligned}
$$

where $dW_t \, dZ_t = \rho \, dt$.

- fSABR is the natural fBm extension of the SABR model of [Hagan et al.] (with $\beta = 1$).
- If $H < 1/2$, the variance of volatility grows sublinearly generating a natural term structure of ATM volatility skew.
- The fSABR stochastic volatility model is non-Markovian, just like the fractional Bergomi model.

The fractional Stein and Stein (fSS) model

As an even more tractable alternative, consider the following model:

$$\begin{aligned}\frac{dS_t}{S_t} &= \sigma_t\, dZ_t\\ \sigma_t &= \sigma_0 + \eta\, W_t^H.\end{aligned}$$

again with $dW_t\, dZ_t = \rho\, dt$.

- In the fSS model, the volatility σ_t is a normal random variable.
- Again, if $H < 1/2$, the variance of volatility grows sublinearly thus volatility is "mean-reverting".
- The fSS model can be considered a toy version of the more realistic fSABR model where more quantities of interest are explicitly computable.

Simulation of fSS and fSABR models

- First, for each Monte Carlo path, generate the correlated Brownian increments ΔW_t and ΔZ_t.
- Given the ΔW_t, the W_t^H are constructed, for example using the Cholesky decomposition method.

SPX smiles in the fSABR model

- In Figure 9, we show how the fSABR model generates very good fits to the SPX option market as of 04-Feb-2010, a day when the ATM volatility term structure happened to be flat.
- fSABR parameters were: $\bar{\sigma} = 0.235$, $\eta = 0.2/0.235$, $H = 1/4$, $\rho = -0.7$.
  - Note in particular that we have obtained a good fit to the whole volatility surface using a model with very few parameters!
  - Moreover, H and η can be fixed from the VIX market.

fSABR fits to SPX smiles as of 04-Feb-2010

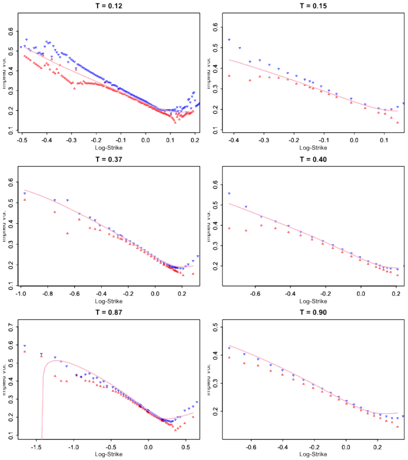

Figure 9 : Red and blue points represent bid and offer SPX implied volatilities; orange smiles are the fSABR fit.

The term structure of ATM VIX volatility

- Consider the following approximation to ATM VIX volatility:

$$\sigma_{\mathrm{BS}}^{VIX}(t) \approx \frac{\eta}{\sqrt{t}} \sqrt{(t+\epsilon)^{2H} - \frac{\epsilon^{2H}}{c_H^2\, 2H}} \tag{13}$$

where $\epsilon = \Delta/2$.

- Up to a constant factor, the approximation (13) is in practice very accurate.
- The approximation (13) is useful for two reasons:
  1. It gives intuition for how the market implied value of H is fixed by the term structure of ATM VIX volatility.
  2. It is much easier to determine H using (13) than from calibration using Monte Carlo.

Numerical results

- Calibrating H to the observed VIX term structure as of 04-Feb-2010 using (13) gave $H = 0.212$.
- We see clearly that H is indeed fixed by the term structure of ATM VIX volatility (or possibly by some other measure of volatility of volatility).
- With $H = 1/4$ and $\eta = 0.52$, we obtain the fit to the VIX option smiles displayed in Figure 10.

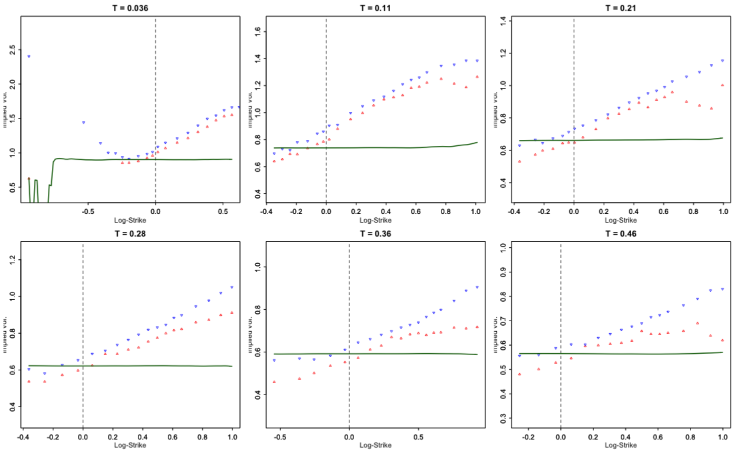

Figure 10 : Red and blue points represent bid and offer VIX implied volatilities as of 04-Feb-2010; Dark green smiles are generated from Monte Carlo simulation of the fSABR model. fSABR parameters were: $\bar{\sigma} = 0.235$, $\eta = 0.52$, $H = 1/4$.

Fit to SPX again

- With H and η fixed by the VIX market, we only have ρ left to achieve a fit to the SPX market.
- With such a low volatility of volatility (relative to $\eta = 0.85$ from the original fit shown in Figure 9), we are forced to set $\rho = -0.99$ which is close to its negative limit.
- We obtain the fit to the SPX market shown in Figure 11.
    - Obviously not as nice as Figure 9!

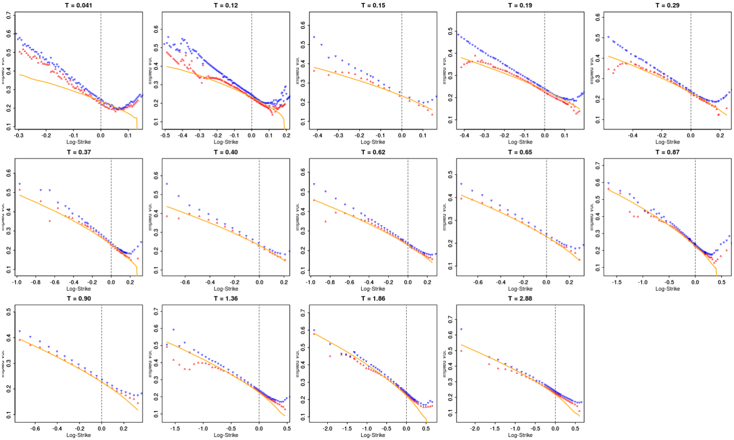

Figure 11 : Red and blue points represent bid and offer SPX implied volatilities as of 04-Feb-2010; Orange smiles are generated from Monte Carlo simulation of the fSABR model. fSABR parameters were: $\bar{\sigma} = 0.235$, $\eta = 0.52$, $H = 1/4$, $\rho = -0.99$.

Summary

- Variance of the log of realized variance exhibits clear power-law scaling.
    - Long memory of RV is rejected by the data.
- The resulting data-driven model (*fBergomi*) is a non-Markovian generalization of the Bergomi model.
    - The conventional Markovian Bergomi market model can be viewed as an accurate approximation to fBergomi.
- The fSABR model, a natural generalization of the SABR model, is a tractable approximation to the fBergomi model.
- fSABR fits observed smiles and skews remarkably well.
    - The value of the Hurst exponent H is fixed by the term structure of VIX at-the-money implied volatility.

References

Torben G Andersen, Tim Bollerslev, Francis X Diebold, and Heiko Ebens, The distribution of realized stock return volatility, *Journal of Financial Economics* **61** (1) 43–76 (2001).

Lorenzo Bergomi, Smile dynamics II, *Risk Magazine* 67–73 (October 2005).

Lorenzo Bergomi and Julien Guyon, The smile in stochastic volatility models, *SSRN* (2011).

Fabienne Comte, Laure Coutin and Eric Renault, Affine fractional stochastic volatility models, *Annals of Finance* **8**(2–3), 337–378 (2012).

Fabienne Comte and Eric Renault, Long memory continuous time models, *Journal of Econometrics* **73**(1) 101–149 (1996).

Fabienne Comte and Eric Renault, Long memory in continuous-time stochastic volatility models, *Mathematical Finance* **8** 29–323(1998).

Laurent Decreusefond, Stochastic integration with respect to Volterra processes, *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* **41**(2) 123–149 (2005).

Jim Gatheral, *The Volatility Surface: A Practitioner's Guide*, Wiley Finance (2006).

P Hagan, D Kumar, A Lesniewski, and D Woodward, Managing smile risk, *Wilmott Magazine*, 84–108, (September 2002).

Rui Vilela Mendes and Maria João Oliveira,, A data-reconstructed fractional volatility model, *arXiv* available at `http://arxiv.org/abs/math/0602013v2` (2006).