Rough volatility

Jim Gatheral
(joint work with Christian Bayer, Peter Friz,
Thibault Jaisson, Andrew Lesniewski, and Mathieu Rosenbaum)

Baruch College
The City University of New York

National School of Development, Peking University,
Tuesday November 10, 2015

Outline of this talk

- The volatility surface: Stylized facts
- A remarkable monofractal scaling property of historical volatility
- Fractional Brownian motion (fBm)
- The Rough Fractional Stochastic Volatility (RFSV) model
- The Rough Bergomi (rBergomi) model
- Fits to SPX
- Forecasting the variance swap curve

SPX volatility smiles as of 15-Sep-2005

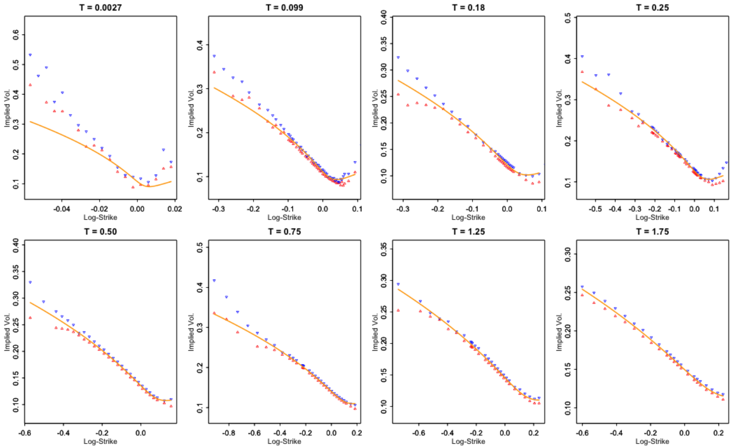

Figure 1: SVI fit superimposed on smiles.

The SPX volatility surface as of 15-Sep-2005

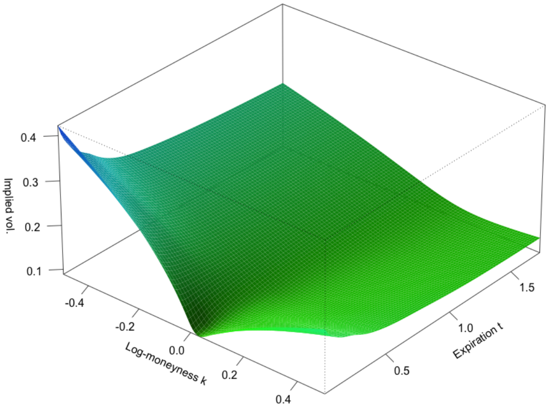

Figure 2: The SPX volatility surface as of 15-Sep-2005 (Figure 3.2 of The Volatility Surface).

Interpreting the smile

- We could say that the volatility smile (at least in equity markets) reflects two basic observations:
  - Volatility tends to increase when the underlying price falls,
    - hence the negative skew.
  - We don't know in advance what realized volatility will be,
    - hence implied volatility is increasing in the wings.
- It's implicit in the above that more or less any model that is consistent with these two observations will be able to fit one given smile.
  - Fitting two or more smiles simultaneously is much harder.
    - Heston for example fits a maximum of two smiles simultaneously.
    - SABR can only fit one smile at a time.

Term structure of at-the-money skew

- What really distinguishes between models is how the generated smile depends on time to expiration.
  - In particular, their predictions for the term structure of ATM volatility skew defined as

$$\psi(\tau) := \left| \frac{\partial}{\partial k} \sigma_{\text{BS}}(k,\tau) \right|_{k=0}.$$

Term structure of SPX ATM skew as of 15-Sep-2005

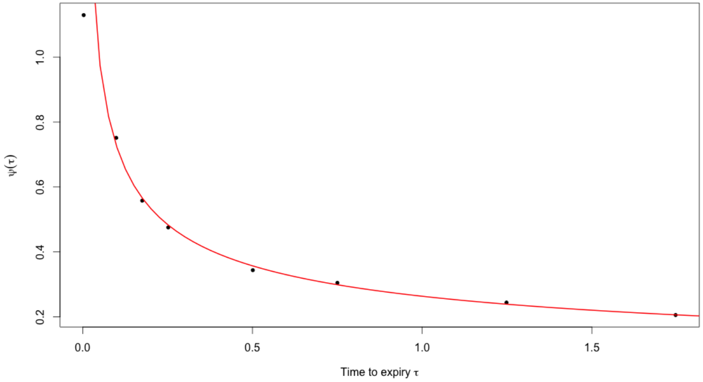

Figure 3: Term structure of ATM skew as of 15-Sep-2005, with power law fit $\tau^{-0.44}$ superimposed in red.

Stylized facts

- Although the levels and orientations of the volatility surfaces change over time, their rough shape stays very much the same.
  - It's then natural to look for a time-homogeneous model.
- The term structure of ATM volatility skew

$$\psi(\tau) \sim \frac{1}{\tau^{\alpha}}$$

with $\alpha \in (0.3, 0.5)$.

Motivation for Rough Volatility 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.

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

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

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

$$C^{x\,\xi} = \int_0^T dt \int_t^T du\, \frac{\mathbb{E}\left[dx_t\, d\xi_t(u)\right]}{dt}. \qquad (2)$$

  - Thus, given a stochastic model, defined in terms of an SDE, we can easily (at least in principle) compute this smile approximation.

The Bergomi model

- The n-factor Bergomi variance curve model reads:

$$\xi_t(u) = \xi_0(u) \exp\left\{ \sum_{i=1}^{n} \eta_i \int_0^t e^{-\kappa_i (t-s)} \, dW_s^{(i)} + \text{ drift } \right\}. \tag{3}$$

- 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.

ATM skew in the Bergomi model

- The Bergomi model generates a term structure of volatility skew $\psi(\tau)$ that is something like

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

  - This functional form is related to the term structure of the autocorrelation function.
  - Which is in turn driven by the exponential kernel in the exponent in (3).
- The observed $\psi(\tau) \sim \tau^{-\alpha}$ for some α.
- It's tempting to replace the exponential kernels in (3) with a power-law kernel.

Tinkering with the Bergomi model

- This would give a model of the form

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

which looks similar to

$$\xi_t(u) = \xi_0(u) \, \exp\left\{ \eta \, W_t^H + \text{ drift } \right\}$$

where W_t^H is fractional Brownian motion.

Motivation for Rough Volatility II: Power-law scaling of the volatility process

- 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

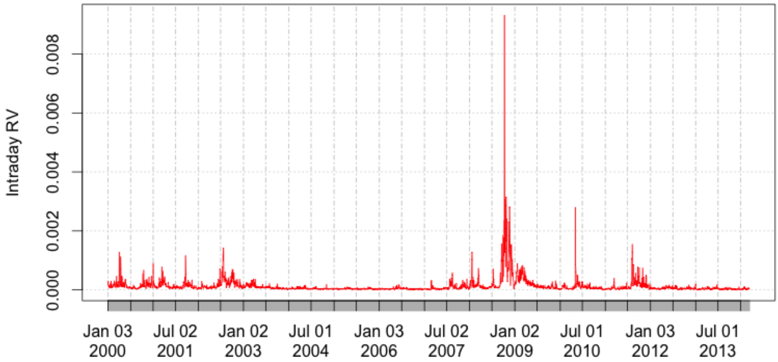

Figure 4: KRV estimates of SPX realized variance from 2000 to 2014.

The smoothness of the volatility process

- For $q \geq 0$, we define the qth sample moment of differences of log-volatility at a given lag Δ[1]:

$$m(q, \Delta) = \langle |\log \sigma_{t+\Delta} - \log \sigma_t|^q \rangle$$

- For example

$$m(2, \Delta) = \langle (\log \sigma_{t+\Delta} - \log \sigma_t)^2 \rangle$$

is just the sample variance of differences in log-volatility at the lag Δ.

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

Scaling of $m(q, \Delta)$ with lag Δ

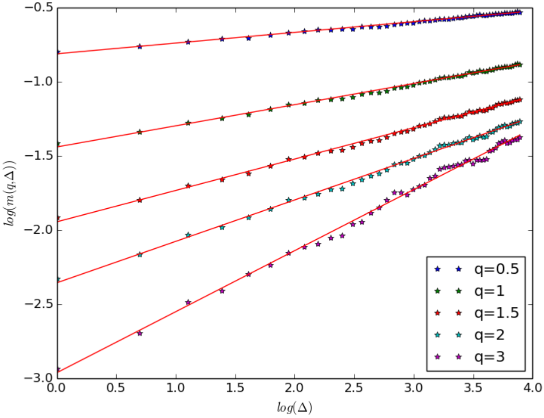

Figure 5: $\log m(q, \Delta)$ as a function of $\log \Delta$, SPX.

Monofractal scaling result

- From the log-log plot Figure 5, we see that for each q, $m(q, \Delta) \propto \Delta^{\zeta_q}$.
- Furthermore, we find the monofractal scaling relationship

$$\zeta_q = q\,H$$

with $H \approx 0.14$.

  - Note however that H does vary over time, in a narrow range.
  - Note also that our estimate of H is biased high because we proxied instantaneous variance v_t with its average over each day $\frac{1}{T}\int_0^T v_t\,dt$, where T is one day.

Distributions of $(\log \sigma_{t+\Delta} - \log \sigma_t)$ for various lags Δ

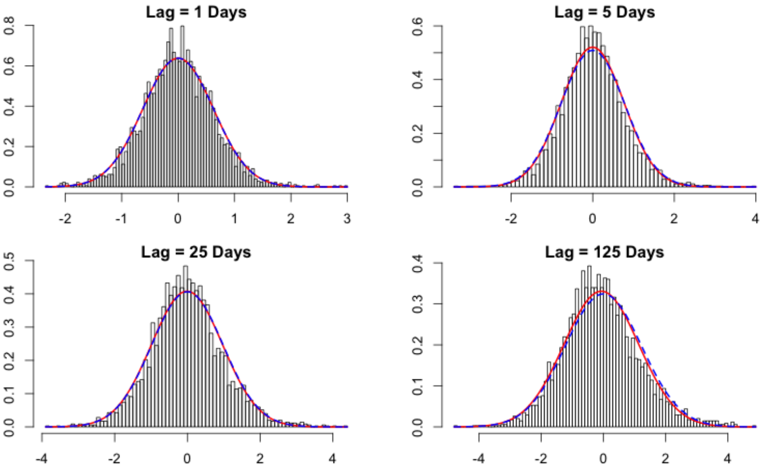

Figure 6: Histograms of $(\log \sigma_{t+\Delta} - \log \sigma_t)$ for various lags Δ; normal fit in red; $\Delta = 1$ normal fit scaled by $\Delta^{0.14}$ in blue.

Estimated H for all indices

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

Index	$\zeta_{0.5}/0.5$	ζ_1	$\zeta_{1.5}/1.5$	$\zeta_2/2$	$\zeta_3/3$
SPX2.rv	0.128	0.126	0.125	0.124	0.124
FTSE2.rv	0.132	0.132	0.132	0.131	0.127
N2252.rv	0.131	0.131	0.132	0.132	0.133
GDAXI2.rv	0.141	0.139	0.138	0.136	0.132
RUT2.rv	0.117	0.115	0.113	0.111	0.108
AORD2.rv	0.072	0.073	0.074	0.075	0.077
DJI2.rv	0.117	0.116	0.115	0.114	0.113
IXIC2.rv	0.131	0.133	0.134	0.135	0.137
FCHI2.rv	0.143	0.143	0.142	0.141	0.138
HSI2.rv	0.079	0.079	0.079	0.080	0.082
KS11.rv	0.133	0.133	0.134	0.134	0.132
AEX.rv	0.145	0.147	0.149	0.149	0.149
SSMI.rv	0.149	0.153	0.156	0.158	0.158
IBEX2.rv	0.138	0.138	0.137	0.136	0.133
NSEI.rv	0.119	0.117	0.114	0.111	0.102
MXX.rv	0.077	0.077	0.076	0.075	0.071
BVSP.rv	0.118	0.118	0.119	0.120	0.120
GSPTSE.rv	0.106	0.104	0.103	0.102	0.101
STOXX50E.rv	0.139	0.135	0.130	0.123	0.101
FTSTI.rv	0.111	0.112	0.113	0.113	0.112
FTSEMIB.rv	0.130	0.132	0.133	0.134	0.134

Table 1: Estimates of ζ_q for all indices in the Oxford-Man dataset.

Universality?

- [Gatheral, Jaisson and Rosenbaum] compute daily realized variance estimates over one hour windows for DAX and Bund futures contracts, finding similar scaling relationships.
- We have also checked that Gold and Crude Oil futures scale similarly.
    - Although the increments $(\log \sigma_{t+\Delta} - \log \sigma_t)$ seem to be fatter tailed than Gaussian.

A natural model of realized volatility

- Distributions of differences in the log of realized volatility are close to Gaussian.
  - This motivates us to model σ_t as a lognormal random variable.
- Moreover, the scaling property of variance of RV differences suggests the model:

$$\log \sigma_{t+\Delta} - \log \sigma_t = \nu \left(W^H_{t+\Delta} - W^H_t \right) \tag{4}$$

where W^H is fractional Brownian motion.
- In [Gatheral, Jaisson and Rosenbaum], we refer to a stationary version of (4) as the RFSV (for Rough Fractional Stochastic Volatility) model.

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

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

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,

$$\begin{aligned} \frac{dS_t}{S_t} &= \sigma_t \, dZ_t \\ d \log \sigma_t &= -\alpha \left(\log \sigma_t - \theta \right) dt + \gamma \, d\hat{W}_t^H \end{aligned} \qquad (5)$$

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.

RFSV and FSV

- The model (4):

$$\log \sigma_{t+\Delta} - \log \sigma_t = \nu \left(W^H_{t+\Delta} - W^H_t \right) \tag{6}$$

  is not stationary.
  - Stationarity is desirable both for mathematical tractability and also to ensure reasonableness of the model at very large times.
- The RFSV model (the stationary version of (4)) is formally identical to the FSV model. Except that
  - $H < 1/2$ in RFSV vs $H > 1/2$ in FSV.
  - $\alpha\, T \gg 1$ in RFSV vs $\alpha\, T \sim 1$ in FSV

  where T is a typical timescale of interest.

FSV and long memory

- Why did [Comte and Renault] choose $H > 1/2$?
  - Because it has been a widely-accepted stylized fact that the volatility time series exhibits long memory.
- In this technical sense, *long memory* means that the autocorrelation function of volatility decays as a power-law.
- One of the influential papers that established this was [Andersen et al.] which estimated the degree d of fractional integration from daily realized variance data for the 30 DJIA stocks.
  - Using the GPH estimator, they found d around 0.35 which implies that the ACF $\rho(\tau) \sim \tau^{2\,d-1} = \tau^{-0.3}$ as $\tau \to \infty$.
- But every statistical estimator assumes the validity of some underlying model!
  - In the RFSV model,

$$\rho(\Delta) \sim \exp\left\{-\frac{\eta^2}{2}\,\Delta^{2\,H}\right\}.$$

Correlogram and test of scaling

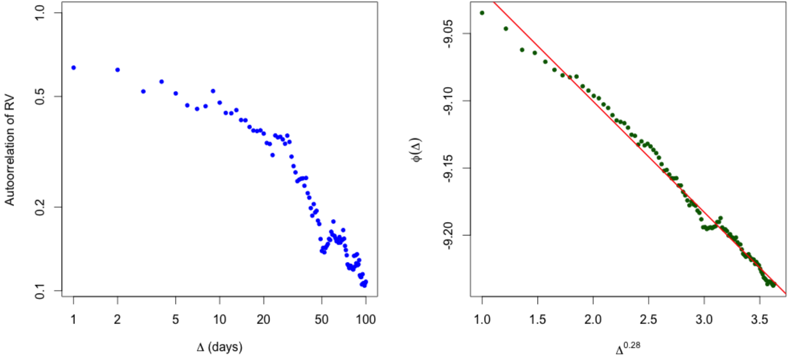

Figure 7: The LH plot is a conventional correlogram of RV; the RH plot is of $\phi(\Delta) := \langle \log\left(\mathrm{cov}(\sigma_{t+\Delta}, \sigma_t) + \langle \sigma_t \rangle^2\right) \rangle$ vs Δ^{2H} with $H = 0.14$. The RH plot again supports the scaling relationship $m(2, \Delta) \propto \Delta^{2H}$.

Model vs empirical autocorrelation functions

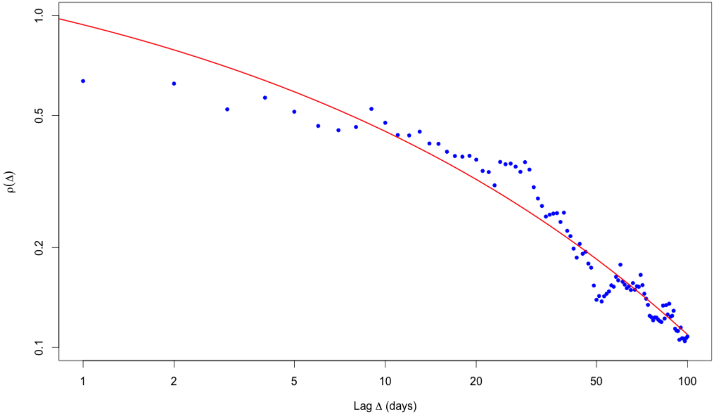

Figure 8: Here we superimpose the RFSV functional form $\rho(\Delta) \sim \exp\left\{-\frac{\eta^2}{2}\,\Delta^{2\,H}\right\}$ (in red) on the empirical curve (in blue).

Volatility is not long memory

- It's clear from Figures 7 and 8 that volatility is not long memory.
- Moreover, the RFSV model reproduces the observed autocorrelation function very closely.
- [Gatheral, Jaisson and Rosenbaum] further simulate volatility in the RFSV model and apply standard estimators to the simulated data.
- Real data and simulated data generate very similar plots and similar estimates of the long memory parameter to those found in the prior literature.
- The RSFV model does not have the long memory property.
  - Classical estimation procedures seem to identify spurious long memory of volatility.

Incompatibility of FSV with realized variance (RV) data

- In Figure 9, we demonstrate graphically that long memory volatility models such as FSV with $H > 1/2$ are not compatible with the RV data.
- In the FSV model, the autocorrelation function $\rho(\Delta) \propto \Delta^{2H-2}$. Then, for long memory, we must have $1/2 < H < 1$.
  - For $\Delta \gg 1/\alpha$, stationarity kicks in and $m(2,\Delta)$ tends to a constant as $\Delta \to \infty$.
  - For $\Delta \ll 1/\alpha$, mean reversion is not significant and $m(2,\Delta) \propto \Delta^{2H}$.

Incompatibility of FSV with RV data

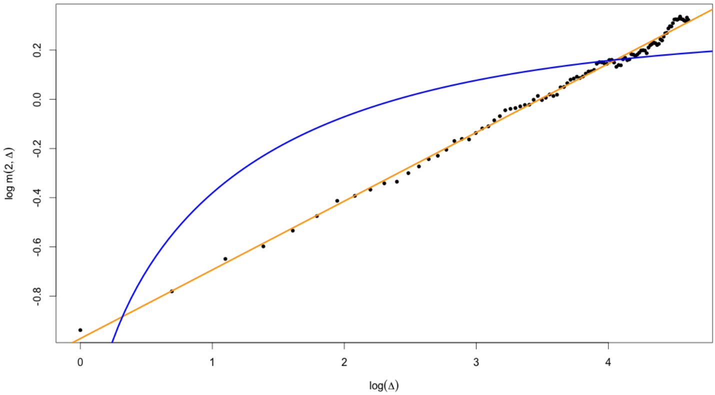

Figure 9: Black points are empirical estimates of $m(2, \Delta)$; the blue line is the FSV model with $\alpha = 0.5$ and $H = 0.53$; the orange line is the RFSV model with $\alpha = 0$ and $H = 0.14$.

Does simulated RSFV data look real?

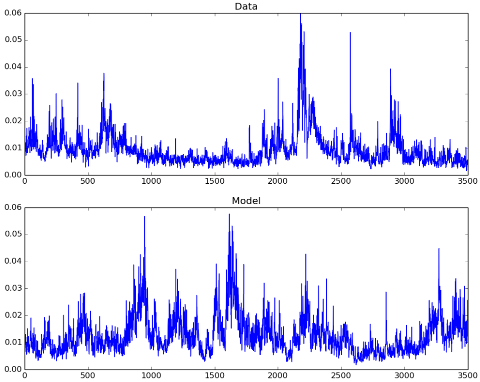

Figure 10: Volatility of SPX (above) and of the RFSV model (below).

Remarks on the comparison

- The simulated and actual graphs look very alike.
  - Persistent periods of high volatility alternate with low volatility periods.
- $H \sim 0.1$ generates very rough looking sample paths (compared with $H = 1/2$ for Brownian motion).
  - Hence *rough volatility*.
- On closer inspection, we observe fractal-type behavior.
  - The graph of volatility over a small time period looks like the same graph over a much longer time period.
- This feature of volatility has been investigated both empirically and theoretically in, for example, [Bacry and Muzy].
  - In particular, their Multifractal Random Walk (MRW) is related to a limiting case of the RSFV model as $H \to 0$.

Pricing under rough volatility

The foregoing behavior suggest the following model for volatility under the real (or historical or physical) measure $\mathbb{P}$:

$$\log \sigma_t = \nu \, W_t^H.$$

Let $\gamma = \frac{1}{2} - H$. We choose the Mandelbrot-Van Ness representation of fractional Brownian motion W^H as follows:

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

where the choice

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

ensures that

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

Pricing under rough volatility

Then

$$
\begin{aligned}
& \log v_u - \log v_t \\
= \; & \nu \, C_H \left\{ \int_t^u \frac{1}{(u-s)^\gamma} \, dW_s^{\mathbb{P}} + \int_{-\infty}^t \left[\frac{1}{(u-s)^\gamma} - \frac{1}{(t-s)^\gamma} \right] dW_s^{\mathbb{P}} \right\} \\
=: \; & 2 \, \nu \, C_H \left[M_t(u) + Z_t(u) \right]. \qquad (7)
\end{aligned}
$$

- Note that $\mathbb{E}^{\mathbb{P}}\left[M_t(u) | \, \mathcal{F}_t \right] = 0$ and $Z_t(u)$ is $\mathcal{F}_t$-measurable.
- To price options, it would seem that we would need to know $\mathcal{F}_t$, the entire history of the Brownian motion W_s for $s < t$!

Pricing under $\mathbb{P}$

Let

$$\tilde{W}_t^{\mathbb{P}}(u) := \sqrt{2\,H} \int_t^u \frac{dW_s^{\mathbb{P}}}{(u-s)^{\gamma}}$$

With $\eta := 2\,\nu\,C_H/\sqrt{2\,H}$ we have $2\,\nu\,C_H\,M_t(u) = \eta\,\tilde{W}_t^{\mathbb{P}}(u)$ so denoting the stochastic exponential by $\mathcal{E}(\cdot)$, we may write

$$\begin{aligned} v_u &= v_t \exp\left\{\eta \tilde{W}_t^{\mathbb{P}}(u) + 2\,\nu\,C_H\,Z_t(u)\right\} \\ &= \mathbb{E}^{\mathbb{P}}\left[v_u|\,\mathcal{F}_t\right]\mathcal{E}\left(\eta\,\tilde{W}_t^{\mathbb{P}}(u)\right). \end{aligned} \tag{8}$$

- The conditional distribution of v_u depends on $\mathcal{F}_t$ only through the variance forecasts $\mathbb{E}^{\mathbb{P}}\left[v_u|\,\mathcal{F}_t\right]$,
- To price options, one does not need to know $\mathcal{F}_t$, the entire history of the Brownian motion $W_s^{\mathbb{P}}$ for $s < t$.

Pricing under $\mathbb{Q}$

Our model under $\mathbb{P}$ reads:

$$v_u = \mathbb{E}^{\mathbb{P}}\left[v_u \middle| \mathcal{F}_t \right] \mathcal{E}\left(\eta \, \tilde{W}_t^{\mathbb{P}}(u) \right). \tag{9}$$

Consider some general change of measure

$$dW_s^{\mathbb{P}} = dW_s^{\mathbb{Q}} + \lambda_s \, ds,$$

where $\{\lambda_s : s > t\}$ has a natural interpretation as the price of volatility risk. We may then rewrite (9) as

$$v_u = \mathbb{E}^{\mathbb{P}}\left[v_u \middle| \mathcal{F}_t \right] \mathcal{E}\left(\eta \, \tilde{W}_t^{\mathbb{Q}}(u) \right) \exp\left\{ \eta \sqrt{2H} \int_t^u \frac{\lambda_s}{(u-s)^{\gamma}} \, ds \right\}.$$

- Although the conditional distribution of v_u under $\mathbb{P}$ is lognormal, it will not be lognormal in general under $\mathbb{Q}$.
  - The upward sloping smile in VIX options means λ_s cannot be deterministic in this picture.

The rough Bergomi (rBergomi) model

Let's nevertheless consider the simplest change of measure

$$dW_s^{\mathbb{P}} = dW_s^{\mathbb{Q}} + \lambda(s)\, ds,$$

where $\lambda(s)$ is a deterministic function of s. Then from (38), we would have

$$\begin{aligned} v_u &= \mathbb{E}^{\mathbb{P}}\left[v_u \middle| \mathcal{F}_t \right] \mathcal{E}\left(\eta\, \tilde{W}_t^{\mathbb{Q}}(u) \right) \exp\left\{ \eta \sqrt{2\,H} \int_t^u \frac{1}{(u-s)^{\gamma}}\, \lambda(s)\, ds \right\} \\ &= \xi_t(u)\, \mathcal{E}\left(\eta\, \tilde{W}_t^{\mathbb{Q}}(u) \right) \end{aligned} \tag{10}$$

where the forward variances $\xi_t(u) = \mathbb{E}^{\mathbb{Q}}\left[v_u \middle| \mathcal{F}_t \right]$ are (at least in principle) tradable and observed in the market.

- $\xi_t(u)$ is the product of two terms:
  - $\mathbb{E}^{\mathbb{P}}\left[v_u \middle| \mathcal{F}_t \right]$ which depends on the historical path $\{W_s, s < t\}$ of the Brownian motion
  - a term which depends on the price of risk $\lambda(s)$.

Features of the rough Bergomi model

- The rBergomi model is a non-Markovian generalization of the Bergomi model:

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

- The rBergomi model is Markovian in the (infinite-dimensional) state vector $\mathbb{E}^{\mathbb{Q}}\left[v_u \middle| \mathcal{F}_t\right] = \xi_t(u)$.
- We have achieved our aim of replacing the exponential kernels in the Bergomi model (3) with a power-law kernel.
  - We may therefore expect that the rBergomi model will generate a realistic term structure of ATM volatility skew.

The stock price process

- The observed anticorrelation between price moves and volatility moves may be modeled naturally by anticorrelating the Brownian motion W that drives the volatility process with the Brownian motion driving the price process.
- Thus
$$\frac{dS_t}{S_t} = \sqrt{v_t}\, dZ_t$$
with
$$dZ_t = \rho\, dW_t + \sqrt{1-\rho^2}\, dW_t^{\perp}$$
where ρ is the correlation between volatility moves and price moves.

Simulation of the rBergomi model

We simulate the rBergomi model as follows:

- Construct the joint covariance matrix for the Volterra process $\tilde{W}$ and the Brownian motion Z and compute its Cholesky decomposition.
- For each time, generate iid normal random vectors and multiply them by the lower-triangular matrix obtained by the Cholesky decomposition to get a $m \times 2\,n$ matrix of paths of $\tilde{W}$ and Z with the correct joint marginals.
- With these paths held in memory, we may evaluate the expectation under $\mathbb{Q}$ of any payoff of interest.

- This procedure is very slow!
    - Speeding up the simulation is work in progress.

Guessing rBergomi model parameters

- The rBergomi model has only three parameters: H, η and ρ.
- If we had a fast simulation, we could just iterate on these parameters to find the best fit to observed option prices. But we don't.
- However, the model parameters H, η and ρ have very direct interpretations:
  - H controls the decay of ATM skew $\psi(\tau)$ for very short expirations
  - The product $\rho\,\eta$ sets the level of the ATM skew for longer expirations.
  - Keeping $\rho\,\eta$ constant but decreasing ρ (so as to make it more negative) pushes the minimum of each smile towards higher strikes.
- So we can guess parameters in practice.

Parameter estimation from historical data

- Both the roughness parameter (or Hurst parameter) H and the volatility of volatility η should be the same under $\mathbb{P}$ and $\mathbb{Q}$.
- Earlier, using the Oxford-Man realized variance dataset, we estimated the Hurst parameter $H_{eff} \approx 0.14$ and volatility of volatility $\nu_{eff} \approx 0.3$.
- However, we not observe the instantaneous volatility σ_t, only $\frac{1}{\delta} \int_0^\delta \sigma_t^2 \, dt$ where δ is roughly $3/4$ of a whole day from close to close.
- Using Appendix C of [Gatheral, Jaisson and Rosenbaum], we rescale finding $H \approx 0.05$ and $\nu \approx 1.7$.
- Also, recall that

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

which yields the estimate $\eta \approx 2.5$.

SPX smiles in the rBergomi model

- In Figures 11 and 12, we show how well a rBergomi model simulation with guessed parameters fits the SPX option market as of February 4, 2010, a day when the ATM volatility term structure happened to be pretty flat.
  - rBergomi parameters were: $H = 0.07$, $\eta = 1.9$, $\rho = -0.9$.
  - Only three parameters to get a very good fit to the whole SPX volatility surface!

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

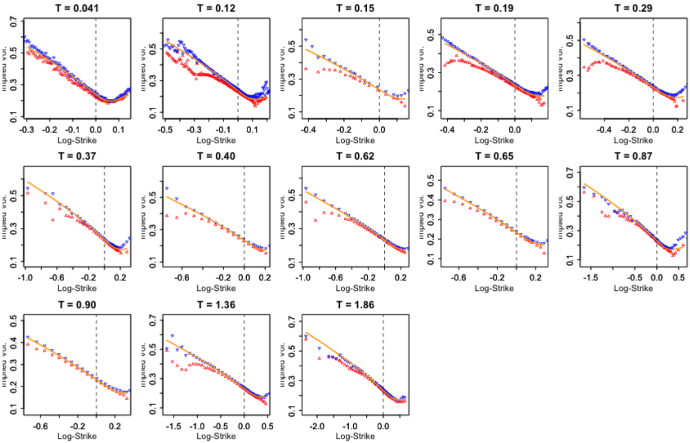

Figure 11: Red and blue points represent bid and offer SPX implied volatilities; orange smiles are from the rBergomi simulation.

Shortest dated smile as of February 4, 2010

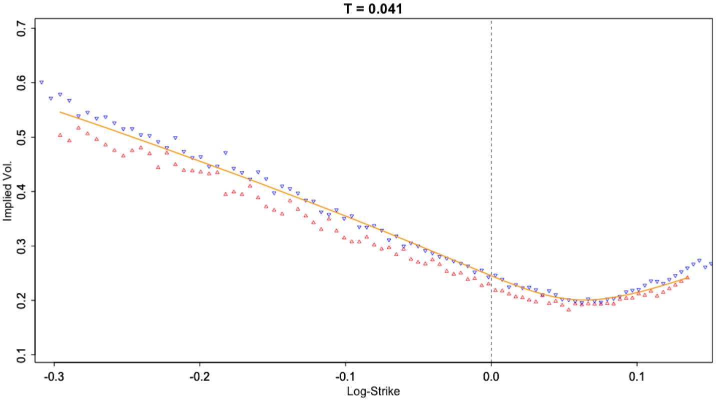

Figure 12: Red and blue points represent bid and offer SPX implied volatilities; orange smile is from the rBergomi simulation.

ATM volatilities and skews

In Figures 13 and 14, we see just how well the rBergomi model can match empirical skews and vols. Recall also that the parameters we used are just guesses!

Term structure of ATM skew as of February 4, 2010

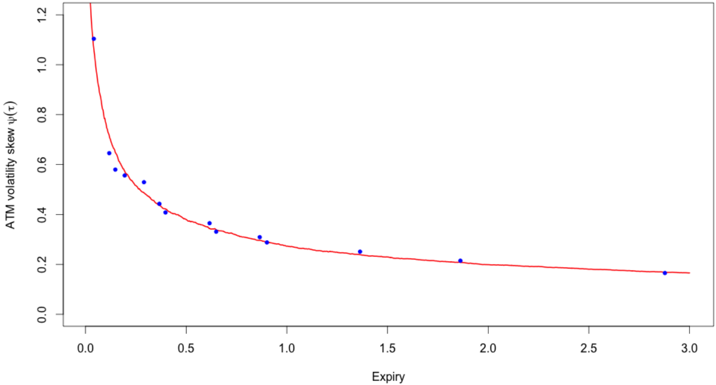

Figure 13: Blue points are empirical skews; the red line is from the rBergomi simulation.

Term structure of ATM vol as of February 4, 2010

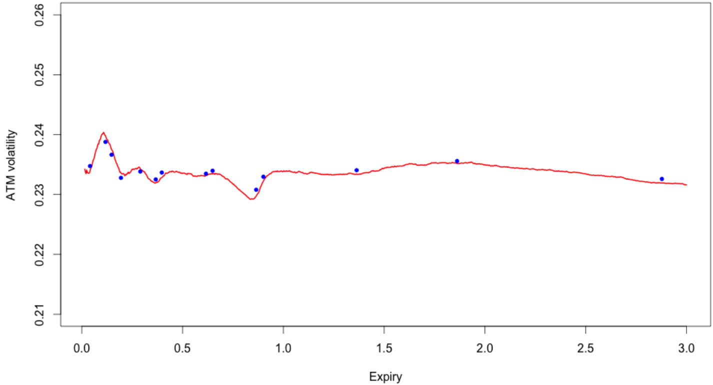

Figure 14: Blue points are empirical ATM volatilities; the red line is from the rBergomi simulation.

Another date

- Now we take a look at another date: August 14, 2013, two days before the last expiration date in our dataset.
  - Options set at the open of August 16, 2013 so only one trading day left.
- Note in particular that the extreme short-dated smile is well reproduced by the rBergomi model.
  - There is no need to add jumps!

SPX smiles as of August 14, 2013

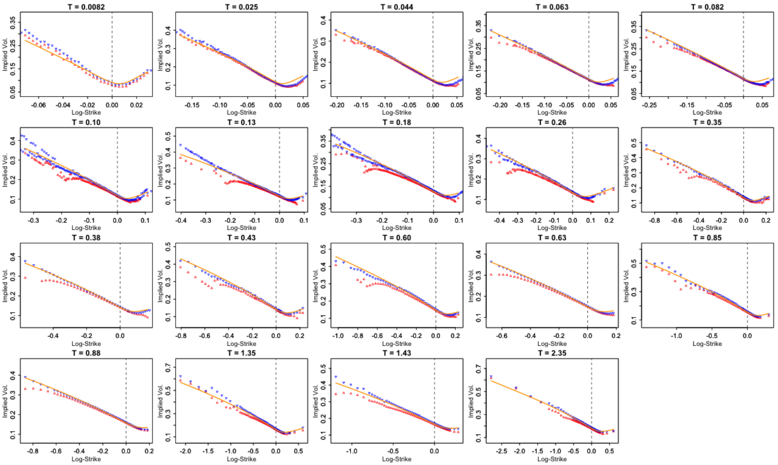

Figure 15: Red and blue points represent bid and offer SPX implied volatilities; orange smiles are from the rBergomi simulation.

The forecast formula

- In the RFSV model (4), $\log v_t \approx 2\,\nu\, W^H_t + C$ for some constant C.
- [Nuzman and Poor] show that $W^H_{t+\Delta}$ is conditionally Gaussian with conditional expectation

$$\mathbb{E}[W^H_{t+\Delta}|\mathcal{F}_t] = \frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\int_{-\infty}^{t}\frac{W^H_s}{(t-s+\Delta)(t-s)^{H+1/2}}ds$$

and conditional variance

$$\text{Var}[W^H_{t+\Delta}|\mathcal{F}_t] = c\,\Delta^{2H}.$$

where

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

The forecast formula

- Thus, we obtain

Variance forecast formula

$$\mathbb{E}^{\mathbb{P}}\left[v_{t+\Delta}\middle|\,\mathcal{F}_t\right]=\exp\left\{\mathbb{E}^{\mathbb{P}}\left[\log(v_{t+\Delta})\middle|\,\mathcal{F}_t\right]+2\,c\,\nu^2\Delta^{2\,H}\right\} \tag{11}$$

where

$$\begin{aligned}&\mathbb{E}^{\mathbb{P}}\left[\log v_{t+\Delta}\middle|\,\mathcal{F}_t\right]\\&=\frac{\cos(H\pi)}{\pi}\Delta^{H+1/2}\int_{-\infty}^{t}\frac{\log v_s}{(t-s+\Delta)(t-s)^{H+1/2}}ds.\end{aligned}$$

Forecasting the variance swap curve

For each of 2,658 days from Jan 27, 2003 to August 31, 2013:

- We compute proxy variance swaps from closing prices of SPX options sourced from OptionMetrics (www.optionmetrics.com) via WRDS.
- We form the forecasts $\mathbb{E}^{\mathbb{P}}\left[v_u | \mathcal{F}_t\right]$ using (11) with 500 lags of SPX RV data sourced from The Oxford-Man Institute of Quantitative Finance (http://realized.oxford-man.ox.ac.uk).
  - We note that the actual variance swap curve is a factor (of roughly 1.4) higher than the forecast, which we may attribute to overnight movements of the index.
  - Forecasts must therefore be rescaled to obtain close-to-close realized variance forecasts.

The RV scaling factor

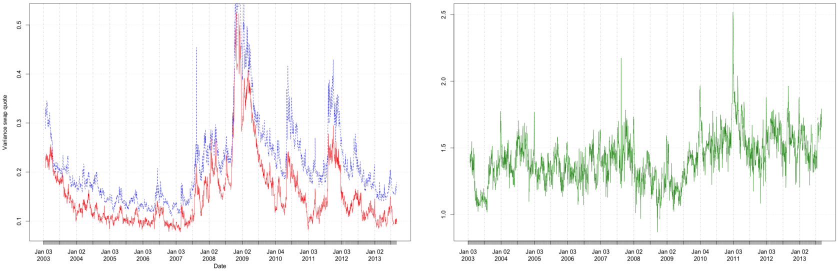

Figure 16: The LH plot shows actual (proxy) 3-month variance swap quotes in blue vs forecast in red (with no scaling factor). The RH plot shows the ratio between 3-month actual variance swap quotes and 3-month forecasts.

The Lehman weekend

- Empirically, it seems that the variance curve is a simple scaling factor times the forecast, but that this scaling factor is time-varying.
- Recall that as of the close on Friday September 12, 2008, it was widely believed that Lehman Brothers would be rescued over the weekend. By Monday morning, we knew that Lehman had failed.
- In Figure 17, we see that variance swap curves just before and just after the collapse of Lehman are just rescaled versions of the RFSV forecast curves.

Actual vs predicted over the Lehman weekend

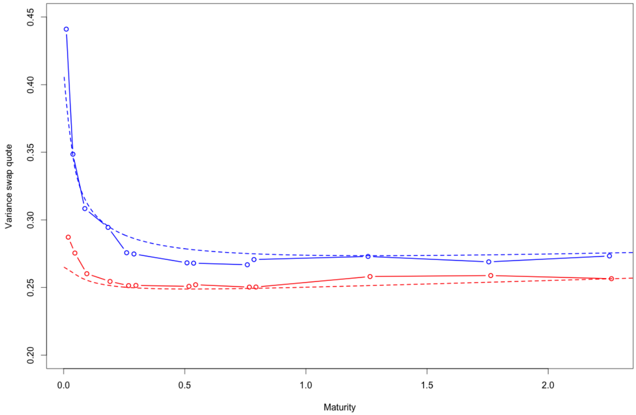

Figure 17: SPX variance swap curves as of September 12, 2008 (red) and September 15, 2008 (blue). The dashed curves are RFSV model forecasts rescaled by the 3-month ratio (1.29) as of the Friday close.

Remarks

We note that

- The actual variance swaps curves are very close to the forecast curves, up to a scaling factor.
- We are able to explain the change in the variance swap curve with only one extra observation: daily variance over the trading day on Monday 15-Sep-2008.
- The SPX options market appears to be backward-looking in a very sophisticated way.

The Flash Crash

- The so-called Flash Crash of Thursday May 6, 2010 caused intraday realized variance to be much higher than normal.
- In Figure 18, we plot the actual variance swap curves as of the Wednesday and Friday market closes together with forecast curves rescaled by the 3-month ratio as of the close on Wednesday May 5 (which was 2.52).
- We see that the actual variance curve as of the close on Friday is consistent with a forecast from the time series of realized variance that *includes* the anomalous price action of Thursday May 6.
- In Figure 19 we see that the actual variance swap curve on Monday, May 10 is consistent with a forecast that excludes the Flash Crash.
    - Volatility traders realized that the Flash Crash should not influence future realized variance projections.

Around the Flash Crash

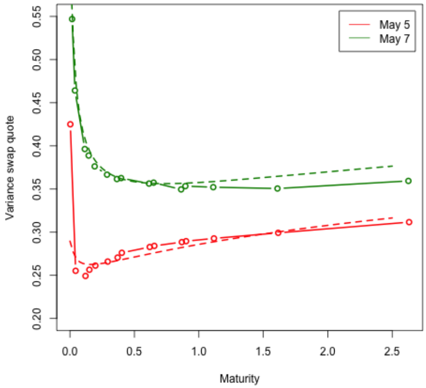

Figure 18: S&P variance swap curves as of May 5, 2010 (red) and May 7, 2010 (green). The dashed curves are RFSV model forecasts rescaled by the 3-month ratio (2.52) as of the close on Wednesday May 5.

The weekend after the Flash Crash

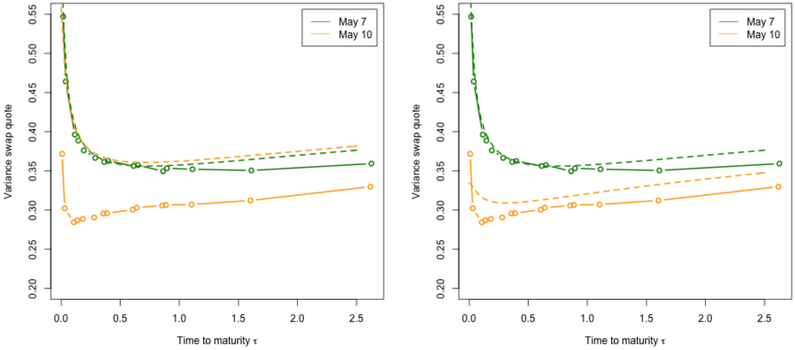

Figure 19: LH plot: The May 10 actual curve is inconsistent with a forecast that includes the Flash Crash. RH plot: The May 10 actual curve is consistent with a forecast that excludes the Flash Crash.

Summary

- We uncovered a remarkable monofractal scaling relationship in historical volatility.
- This leads to a natural non-Markovian stochastic volatility model under $\mathbb{P}$.
- The simplest specification of $\frac{d\mathbb{Q}}{d\mathbb{P}}$ gives a non-Markovian generalization of the Bergomi model.
  - The history of the Brownian motion $\{W_s, s < t\}$ required for pricing is encoded in the forward variance curve, which is observed in the market.
- This model fits the observed volatility surface surprisingly well with very few parameters.
- For perhaps the first time, we have a simple consistent model of historical and 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).
- Christian Bayer, Peter Friz and Jim Gatheral, Pricing under rough volatility, *Quantitative Finance, forthcoming. Available at SSRN 2554754*, (2015).
- Emmanuel Bacry and Jean-François Muzy, Log-infinitely divisible multifractal processes, *Communications in Mathematical Physics* **236**(3) 449–475 (2003).
- Lorenzo Bergomi, Smile dynamics II, *Risk Magazine* 67–73 (October 2005).
- Lorenzo Bergomi and Julien Guyon, Stochastic volatility's orderly smiles. *Risk Magazine* 60–66, (May 2012).
- Fabienne Comte and Eric Renault, Long memory in continuous-time stochastic volatility models, *Mathematical Finance* **8** 29–323(1998).
- Jim Gatheral, *The Volatility Surface: A Practitioner's Guide*, Wiley Finance (2006).
- Jim Gatheral and Antoine Jacquier, Arbitrage-free SVI volatility surfaces, *Quantitative Finance*, **14**(1) 59–71 (2014).
- Jim Gatheral, Thibault Jaisson and Mathieu Rosenbaum, Volatility is rough, *Available at SSRN 2509457*, (2014).
- Carl J. Nuzman and H. Vincent Poor, Linear estimation of self-similar processes via Lamperti's transformation, *Journal of Applied Probability* **37**(2) 429–452 (2000).