



Pricing under rough volatility

Christian Bayer, Peter Friz & Jim Gatheral

To cite this article: Christian Bayer, Peter Friz & Jim Gatheral (2015): Pricing under rough volatility, Quantitative Finance, DOI: [10.1080/14697688.2015.1099717](https://doi.org/10.1080/14697688.2015.1099717)

To link to this article: <http://dx.doi.org/10.1080/14697688.2015.1099717>



Published online: 18 Nov 2015.



Submit your article to this journal [↗](#)



Article views: 7



View related articles [↗](#)



View Crossmark data [↗](#)

Pricing under rough volatility

CHRISTIAN BAYER[†], PETER FRIZ[‡] and JIM GATHERAL^{*§}

[†]Weierstraß-Institut Berlin, Berlin, Germany

[‡]Technische Universität Berlin, Berlin, Germany

[§]Baruch College, City University of New York, New York, USA

(Received 24 January 2015; accepted 3 September 2015; published online 17 November 2015)

From an analysis of the time series of realized variance using recent high-frequency data, Gatheral *et al.* [Volatility is rough, 2014] previously showed that the logarithm of realized variance behaves essentially as a fractional Brownian motion with Hurst exponent H of order 0.1, at any reasonable timescale. The resulting Rough Fractional Stochastic Volatility (RFSV) model is remarkably consistent with financial time series data. We now show how the RFSV model can be used to price claims on both the underlying and integrated variance. We analyse in detail a simple case of this model, the rBergomi model. In particular, we find that the rBergomi model fits the SPX volatility markedly better than conventional Markovian stochastic volatility models, and with fewer parameters. Finally, we show that actual SPX variance swap curves seem to be consistent with model forecasts, with particular dramatic examples from the weekend of the collapse of Lehman Brothers and the Flash Crash.

Keywords: Arbitrage pricing; Volatility surfaces; Stochastic volatility; Fractional Brownian motion; Options pricing; Volatility modelling; Bergomi–Guyon expansion

1. Introduction

From an analysis of the time series of realized variance using recent high-frequency data, Gatheral *et al.* (2014) showed that the logarithm of realized variance behaves essentially as a fractional Brownian motion with Hurst exponent H of order 0.1, at any reasonable timescale. The following stationary Rough Fractional Stochastic Volatility (RFSV) model was proposed:

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu_t dt + \sigma_t dZ_t \\ \sigma_t &= \exp\{X_t\}, \quad t \in [0, T], \end{aligned} \quad (1.1)$$

where μ_t is a suitable drift term, Z_t is a standard Brownian motion, X_t is a fractional Ornstein–Uhlenbeck process (fOU process for short) satisfying

$$dX_t = \nu dW_t^H - \alpha(X_t - m)dt,$$

where $m \in \mathbb{R}$, and ν and α are positive parameters, see Cheridito *et al.* (2003), with Z and W^H correlated in general. Recall that sample paths of fractional Brownian motion W^H are $(H - \varepsilon)$ -Hölder (but not H -Hölder and hence ‘rougher’ than Brownian motion whenever $H < 1/2$). The reversion timescale is understood to be very long so that $\alpha T \ll 1$ for any reasonable timescale T of practical interest (from one day to ten years say), in which case, the log-variance behaves locally (at timescales smaller than T) as a fractional Brownian

motion (fBm). The RFSV model is remarkably consistent with financial time series data. Moreover, the RFSV model has a quantitative market microstructure-based foundation based on the modelling of order flow using Hawkes processes, see Gatheral *et al.* (2014, Section 6) and Jaisson and Rosenbaum (2015).

On the other hand, from the perspective of options pricing, it is well known that conventional low-dimensional Markovian stochastic volatility models such as the Hull and White, Heston, and SABR models generate implied volatility surfaces whose shapes differ substantially from that of the empirically observed volatility surface. A typical such volatility surface generated from a ‘stochastic volatility inspired’ (SVI) (Gatheral and Jacquier 2014) fit to closing SPX option prices as of August 14, 2013[¶] is shown in figure 1. It is a stylized fact that, at least in equity markets, although the level and orientation of the volatility surface do change over time, the general overall shape of the volatility surface does not change, at least to a first approximation. This suggests that it is desirable to model volatility as a time-homogenous process, i.e. a process whose parameters are independent of price and time.

Given an implied volatility smile for a single expiration, little can be said about the process generating it; any process that generates uncertain integrated variance from inception

[¶]Closing prices of SPX options for all available strikes and expirations were sourced from OptionMetrics (www.optionmetrics.com) via Wharton Research Data Services (WRDS).

*Corresponding author. Email: jim.gatheral@baruch.cuny.edu

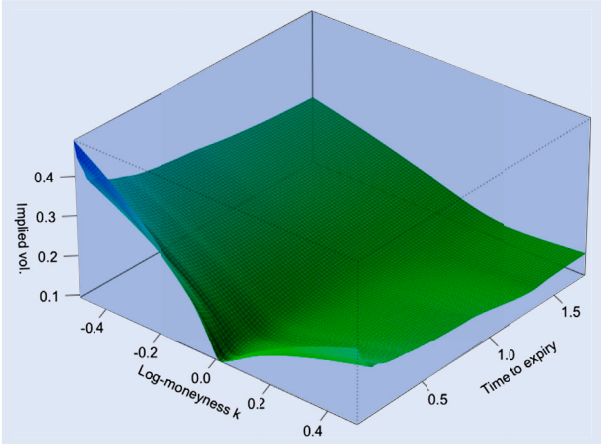


Figure 1. The SPX volatility surface as of August 14, 2013. Time measured in years.

to expiration and with some correlation between changes in volatility and returns of the underlying might suffice. To say more about the underlying process, the scaling of smiles with respect to time to expiration needs to be examined. In particular, one feature of the volatility surface that really does distinguish between models is the term structure of at-the-money (ATM) volatility skew defined as

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

where $\tau = T - t$ denotes time to expiration and $k = \log K/S_t$ is the log-strike. In conventional stochastic volatility models, the ATM volatility skew $\psi(\tau)$ is constant for short dates and inversely proportional to τ for long dates. Empirically, as shown in figure 2, we observe that $\psi(\tau)$ is proportional to $1/\tau^\alpha$ for some $0 < \alpha < 1/2$ over a very wide range of expirations.

Let $v_u = \sigma_u^2$ denote instantaneous variance at time $u > t$ (we refer to appendix 1 for a reminder of the various notions of variance and volatility used in this work). Then the forward variance curve is given by

$$\xi_t(u) = \mathbb{E}[v_u | \mathcal{F}_t], \quad u \geq t.$$

Bergomi and Guyon (2012) derive a small noise expansion for the smile in a stochastic volatility model written in the following forward variance curve form:

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{\xi_t(t)} dZ_t \\ d\xi_t(u) &= \varepsilon \lambda(t, u, \xi_t) dW_t \end{aligned} \quad (1.2)$$

where Z_t is a Brownian motion driving the asset price S_t , W_t is a (suitably correlated) d -dimensional Brownian motion driving the evolution of the forward variance curve, and ε is a dimensionless expansion parameter. To first order in ε , the Bergomi–Guyon expansion takes the form[†]

$$\sigma_{BS}(k, T) = \hat{\sigma}_T + \varepsilon \sqrt{\frac{w}{T}} \frac{1}{2w^2} C^{x\xi} k + O(\varepsilon^2) \quad (1.3)$$

[†]The ATM volatility term $\hat{\sigma}_T$ in the Bergomi–Guyon expansion can also be expressed explicitly to order ε in terms of w and $C^{x\xi}$. See for example equation (7.2).

where the log-strike $k = \log K/S_0$, $w = \int_0^T \xi_0(s) ds$ is total variance to expiration T , and the autocorrelation functional

$$C^{x\xi} = \int_0^T dt \int_t^T du \frac{\mathbb{E}[dx_t d\xi_t(u)]}{dt}. \quad (1.4)$$

where $x_t = \log S_t/S_0$. Thus, given a stochastic model written in the forward variance curve form (1.2), we can easily (at least in principle) compute the term structure of ATM skew $\psi(\tau)$ to first order in ε (which corresponds to first order in volatility of volatility).

One well-known example of a stochastic volatility model expressed in forward variance curve form is the Bergomi model (Bergomi 2005). The n -factor Bergomi variance curve model may be written in the form

$$\xi_t(u) = \xi_0(u) \mathcal{E} \left(\sum_{i=1}^n \eta_i \int_0^t e^{-\kappa_i(u-s)} dW_s^{(i)} \right) \quad (1.5)$$

where $\mathcal{E}(\cdot)$ denotes the stochastic exponential.

Remark 1.1 For a continuous (semi)martingale Z , the stochastic exponential is classically defined as $\mathcal{E}(Z)_t = \exp(Z_t - Z_0 - \frac{1}{2}[Z, Z]_{0,t})$. If Z is a local martingale, then so is $\mathcal{E}(Z)$. On the other hand, for a zero-mean Gaussian random variable Ψ , it will be convenient to introduce the ‘Wick’ exponential as

$$\mathcal{E}(\Psi) = \exp \left(\Psi - \frac{1}{2} \mathbb{E}[\Psi^2] \right).$$

When Ψ is the increment of a Gaussian martingale – such as $\int_0^t f(s) dW_s$ with deterministic integrand – the two notions coincide. Note however that we only ever use the notation \mathcal{E} as shorthand; Wick calculus is not required for any of our arguments.

$\xi_t(u)$ is thus a martingale in t , consistent with the fact that forward variances are tradable. As was pointed out by Bergomi, the entire forward variance curve $\xi_t(\cdot) = \{\xi_t(u) : u > t\}$ is determined by n -factors, each of OU-type. Indeed, in the case $n = 1$ (for notational simplicity only) one has

$$\xi_t(u) = \xi_0(u) \exp \left(\eta e^{-\kappa(u-t)} Y_t - \frac{1}{2} \eta^2 e^{-2\kappa(u-t)} \mathbb{E}[Y_t^2] \right)$$

in terms of a scalar OU process,

$$dY_t = -\kappa Y_t dt + dW_t.$$

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 eight parameters: $\eta_1, \eta_2, \kappa_1, \kappa_2$, and the (constant) correlations $\rho_{Z, W^{(1)}}, \rho_{Z, W^{(2)}}, \rho_{W^{(1)}, W^{(2)}}$, in addition to the initial forward variance curve $\xi_0(u)$. When calibrating the two-factor Bergomi model to option prices, we find that it is already over-parameterized. Any combination of the parameters η_i, κ_i , and the various correlation parameters that gives a roughly $1/\sqrt{T}$ term structure of ATM skew fits well enough. Moreover, the calibrated correlations between the Brownian increments $dW_s^{(i)}$ tend to be high.

The Bergomi model generates a term structure of volatility skew $\psi(\tau)$ that has the qualitative form

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

Indeed, it can be seen from the Bergomi–Guyon expansion that this functional form is related to the term structure of the

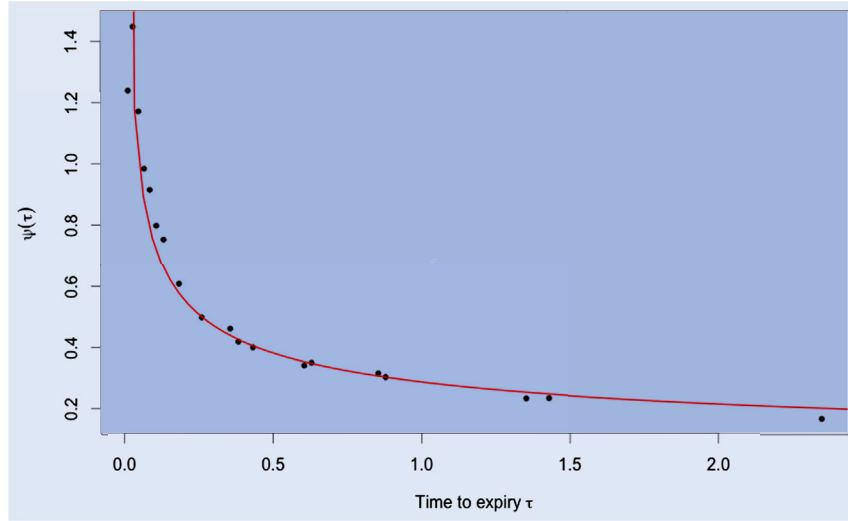


Figure 2. The black dots are non-parametric estimates of the S&P at-the-money (ATM) volatility skews as of August 14, 2013; the red curve is the power-law fit $\psi(\tau) = A\tau^{-0.407}$, τ measured in years.

autocorrelation functional $C^{x\xi}$, as defined in (1.4), which is in turn driven by the exponential kernels in the exponent in (1.5). To generate the empirically observed $\psi(\tau) \sim \tau^{-\alpha}$ for some α , it would be tempting to replace the exponential kernels in (1.5) with a power-law kernel. This would give a model of the form

$$\xi_t(u) = \xi_0(u) \mathcal{E} \left(\eta \int_0^t \frac{dW_s}{(u-s)^\gamma} \right) \quad (1.6)$$

with $\xi_t(u)$ again a martingale in t . Assuming constant $\xi_0(u) \equiv \sigma_0^2$, and with the Wick interpretation of the stochastic exponential, we would have instantaneous stochastic volatility

$$\begin{aligned} \sigma_t = \sqrt{\xi_t(t)} &= \sigma_0 \sqrt{\mathcal{E} \left(\eta \int_0^t \frac{dW_s}{(t-s)^\gamma} \right)} \\ &= \sigma_0 \exp \left\{ \frac{\eta}{2} V_t - \frac{\eta^2}{4} \mathbb{E}[V_t^2] \right\}, \end{aligned}$$

where $V_t = \int_0^t \frac{dW_s}{(t-s)^\gamma}$ is known as ‘Volterra’ fractional Brownian motion with Hurst parameter $H = 1/2 - \gamma$ and has, similar to classical fractional Brownian motion, $(H - \varepsilon)$ -Hölder sample paths. We note a striking resemblance to the RFSV model (1.1). Moreover, by applying his Martingale expansion to a special case of a model originally proposed by Alòs *et al.* (2007), Fukasawa (2011) shows formally that the volatility skew generated by such models has the form

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

for small τ .

In this paper, we show that the RFSV model does indeed lead naturally to a non-Markovian generalization of the Bergomi model, which we call the Rough Bergomi (rBergomi) model. This model fits the observed volatility surface markedly better than conventional Markovian stochastic volatility models, and with fewer parameters.

1.1. Main results and organization of the paper

Our paper is organized as follows. In section 2, we show how the RFSV model leads naturally to an options pricing

model. In section 3, we analyse a special case of this model, the rBergomi model, where the change of measure from \mathbb{P} to \mathbb{Q} is deterministic. In section 4, we show how to simulate the rBergomi model, and in section 5 we show that volatility surfaces generated using the rBergomi model simulation are remarkably consistent with observed ones (at least on the two specific days presented). In section 6, we examine consistency between the rBergomi model and the VIX options market, finding that in general, the rBergomi model is not consistent with the VIX options market. In section 7, we compute coefficients of the Bergomi–Guyon expansion of the rBergomi model up to second order in volatility of volatility; sadly, we find that this asymptotic expansion does not converge with parameters of practical interest. In section 8, we show that the evolution of market variance swap curves is consistent with forecasts obtained from the historical realized variance time series; we examine the cases of the collapse of Lehman Brothers and the Flash Crash in detail. Finally, in section 9, we summarize and conclude. Some more detailed computations are relegated to the appendix.

2. Change of measure

In Gatheral *et al.* (2014), using estimates of realized variance as proxies for daily instantaneous variances, two startlingly simple regularities were uncovered. Firstly, consistent with many prior studies, distributions of increments of the logarithm of realized variance were found to be close to Gaussian. Second and more interestingly, for reasonable timescales of practical interest (from one day to a few years say), the time series of realized variance was found to be consistent with the simple model

$$\log \sigma_{t+\Delta} - \log \sigma_t = v \left(W_{t+\Delta}^H - W_t^H \right) \quad (2.1)$$

where W^H is fractional Brownian motion, which is simply the RFSV model (1.1) with $\alpha = 0$. This relationship was found to hold for all 21 equity indices in the Oxford-Man database,

Bund futures, Crude Oil futures, and Gold futures. Perhaps this feature of the time series of volatility is universal?

Consider the Mandelbrot–Van Ness representation of fractional Brownian motion W^H in terms of Wiener integrals[†]:

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

where $\gamma = 1/2 - H$ and the choice $C_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}$ ensures that

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

Substituting into (2.1) (and in terms of $v_t = \sigma_t^2$), we obtain the following model for the evolution of v_u under the physical measure \mathbb{P} :

$$\begin{aligned} \log v_u - \log v_t &= 2\nu C_H \left\{ \int_{-\infty}^u \frac{dW_s^{\mathbb{P}}}{(u-s)^\gamma} - \int_{-\infty}^t \frac{dW_s^{\mathbb{P}}}{(t-s)^\gamma} \right\} \\ &= 2\nu C_H \left\{ \int_t^u \frac{1}{(u-s)^\gamma} dW_s^{\mathbb{P}} \right. \\ &\quad \left. + \int_{-\infty}^t \left[\frac{1}{(u-s)^\gamma} - \frac{1}{(t-s)^\gamma} \right] dW_s^{\mathbb{P}} \right\} \\ &=: 2\nu C_H \{M_t(u) + Z_t(u)\}. \end{aligned} \quad (2.2)$$

Note that $Z_t(u)$ is \mathcal{F}_t -measurable, whereas $M_t(u)$ is independent of \mathcal{F}_t , and Gaussian with mean zero, and variance $(u-t)^{2H}/(2H)$. We introduce

$$\tilde{W}_t^{\mathbb{P}}(u) := \sqrt{2H} \int_t^u \frac{dW_s^{\mathbb{P}}}{(u-s)^\gamma} \quad (2.3)$$

which has the same properties as $M_t(u)$, only with variance $(u-t)^{2H}$. With $\eta := 2\nu C_H/\sqrt{2H}$ we have $2\nu C_H M_t(u) = \eta \tilde{W}_t^{\mathbb{P}}(u)$ and so

$$\mathbb{E}^{\mathbb{P}}[v_u | \mathcal{F}_t] = v_t \exp \left\{ 2\nu C_H Z_t(u) + \frac{1}{2} \eta^2 \mathbb{E}|\tilde{W}_t^{\mathbb{P}}(u)|^2 \right\}.$$

As a consequence, in terms of the (Wick) stochastic exponential

$$v_u = v_t \exp \left\{ \eta \tilde{W}_t^{\mathbb{P}}(u) + 2\nu C_H Z_t(u) \right\} \quad (2.4a)$$

$$= \mathbb{E}^{\mathbb{P}}[v_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{P}}(u) \right). \quad (2.4b)$$

At this stage, two comments are in order: First, note that the non-Markovian nature of v_u is hidden in (2.4a) in the dependence of $Z_t(u)$ on the full history of $W^{\mathbb{P}}$. However, (2.4b) reveals that the conditional distribution of v_u depends on \mathcal{F}_t only through the instantaneous variance forecasts $\mathbb{E}^{\mathbb{P}}[v_u | \mathcal{F}_t]$, $u > t$.[‡]

Summarizing, we propose the following model under the physical probability measure \mathbb{P} :

$$\frac{dS_u}{S_u} = \mu_u du + \sqrt{v_u} dZ_u^{\mathbb{P}}, \quad (2.5a)$$

$$v_u = v_t \exp \left\{ \eta \tilde{W}_t^{\mathbb{P}}(u) + 2\nu C_H Z_t(u) \right\}, \quad (2.5b)$$

[†]Strictly speaking, this expression is only formal. The rigorous form, as used in the computations below, exploits the cancellation between the integrands as $s \rightarrow -\infty$.

[‡]This is analogous to what happens in Comte, Coutin and Renault Comte et al. (2012) in the context of their fractionally integrated square root model.

defined in terms of two standard Brownian motions $Z^{\mathbb{P}}$ and $W^{\mathbb{P}}$ with (constant) correlation ρ .

Remark 2.1 The process \tilde{W} is an example of a *Brownian semi-stationary process* (see Corcuera et al. (2013) for a nice introduction), that is a process of the form

$$X_t = \int_{-\infty}^t g(t-s) \sigma_s dW_s$$

for some deterministic weight function g and an adapted intermittency process σ . Such processes have been studied extensively; the application of numerical techniques developed in that context to our model remains to be explored.

2.1. Pricing under \mathbb{Q}

We have specified dynamics (2.5) for the underlying under the physical probability measure \mathbb{P} , defined in terms of two standard Brownian motions $Z^{\mathbb{P}}$ and $W^{\mathbb{P}}$ with constant correlation ρ and a suitable drift term $\mu = \mu(u, \omega)$. Pricing of options on the underlying (for example SPX in the foregoing) at (fixed) time t will be based on an equivalent martingale measure; that is, $\mathbb{Q} \sim \mathbb{P}$ on $[t, T]$ such that the discounted asset price process S_t is a martingale under \mathbb{Q} . In our context,

$$\frac{dS_u}{S_u} = \sqrt{v_u} dZ_u^{\mathbb{Q}}, \quad t \leq u \leq T.$$

which is achieved, on some fixed time horizon $[t, T]$, by a Girsanov change of measure,

$$dZ_u^{\mathbb{Q}} = dZ_u^{\mathbb{P}} + \frac{\mu_u}{\sqrt{v_u}} du, \quad t \leq u \leq T.$$

On the other hand, from (2.3), the Volterra-type process $\tilde{W}_t^{\mathbb{P}}$ has been built from $W^{\mathbb{P}}$, a Brownian motion correlated with $Z^{\mathbb{P}}$ in the sense that

$$W^{\mathbb{P}} = \rho Z^{\mathbb{P}} + \bar{\rho} \bar{Z}^{\mathbb{P}}, \quad \rho^2 + \bar{\rho}^2 = 1$$

where $(Z^{\mathbb{P}}, \bar{Z}^{\mathbb{P}})$ is a pair of independent standard Brownian motions. A general change of measure for the second component is then of the form

$$d\bar{Z}_u^{\mathbb{Q}} = d\bar{Z}_u^{\mathbb{P}} + \gamma_u du,$$

where $\gamma = \gamma(u, \omega)$, for $u \in [t, T]$, is a suitable adapted process, called the *market price of volatility risk*. As a result,

$$dW_u^{\mathbb{Q}} = dW_u^{\mathbb{P}} + [\rho \mu_u / \sqrt{v_u} + \bar{\rho} \gamma_u] du, \quad t \leq u \leq T,$$

which we may rewrite as

$$dW_s^{\mathbb{P}} = dW_s^{\mathbb{Q}} + \lambda_s ds. \quad (2.6)$$

To recap, we have a model (2.4b) that accurately mimics the behaviour of realized variance time series data, written under \mathbb{P} :

$$v_u = \mathbb{E}^{\mathbb{P}}[v_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{P}}(u) \right). \quad (2.7)$$

In particular $\mathbb{E}^{\mathbb{P}}[v_u | \mathcal{F}_t]$ is adapted to the filtration generated by $W^{\mathbb{P}}$ (which we assume coincides with the filtration generated by $W^{\mathbb{Q}}$). We may then rewrite (2.7) as

[§]We set interest rates to zero which is without loss of generality, assuming deterministic interest rates.

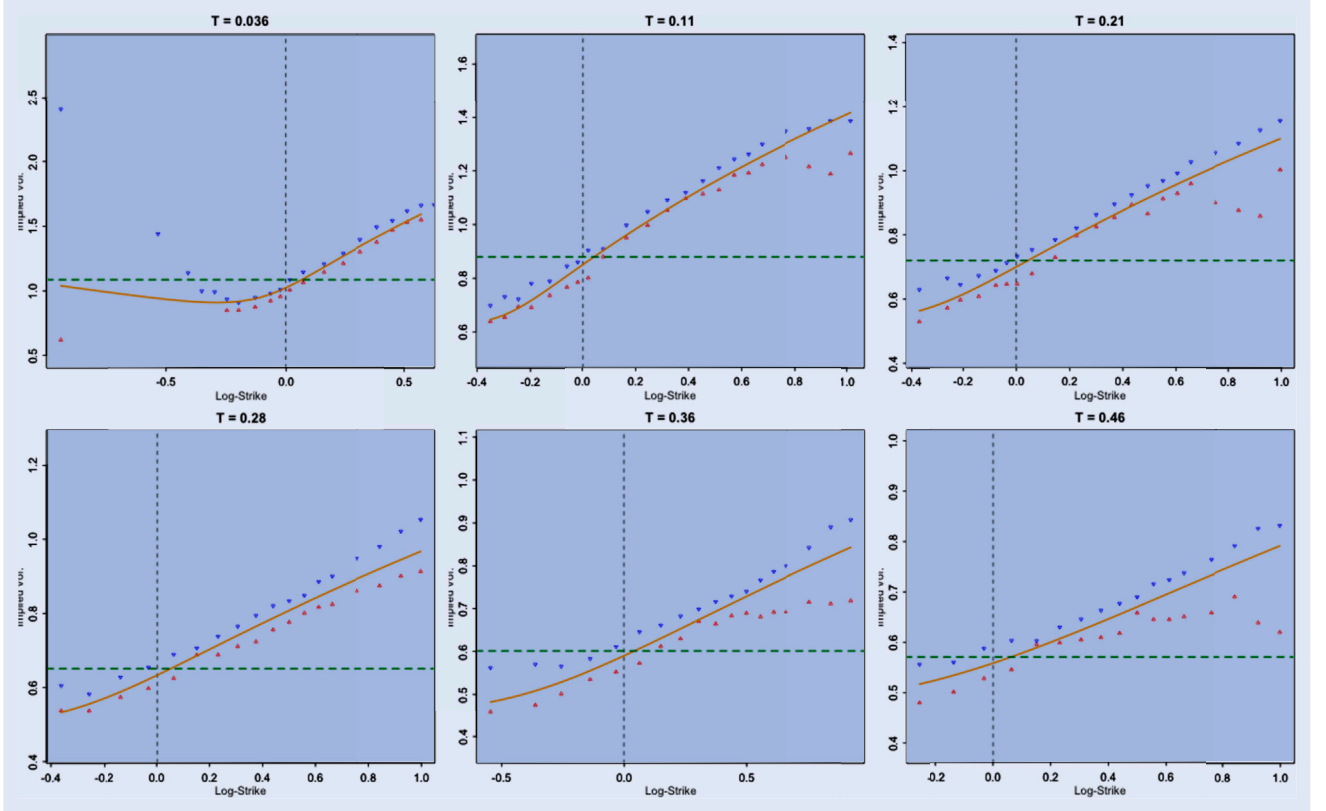


Figure 3. VIX implied volatility smiles as of February 4, 2010. Blue points are ask volatilities; red points are bid volatilities; orange lines are SVI fits; green dashed lines represent the VIX log-strip (VVIX).

$$\begin{aligned}
 v_u &= \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \exp \left\{ \eta \sqrt{2H} \right. \\
 &\quad \times \left. \int_t^u \frac{1}{(u-s)^\gamma} dW_s^{\mathbb{P}} - \frac{\eta^2}{2} (u-t)^{2H} \right\} \\
 &= \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{Q}}(u) \right) \\
 &\quad \times \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{\lambda_s}{(u-s)^\gamma} ds \right\}. \quad (2.8)
 \end{aligned}$$

The last term in the exponent obviously changes the marginal distribution of the v_u ; although the conditional distribution of v_u under \mathbb{P} is lognormal, it will not be lognormal in general under \mathbb{Q} .

VIX smiles and the change of measure

In the case of SPX, it is obvious from the shape of VIX implied volatility smiles that the change of measure cannot be deterministic. If the change of measure were deterministic, it follows from (2.8) that v_u would be conditionally lognormal, VIX would also be approximately lognormal and so the VIX implied volatility smiles would be approximately flat, a well-known problem with the conventional Bergomi model. In contrast, we observe VIX smiles that are strongly upward sloping (see figure 3 for example), reflecting the intuition that high-volatility scenarios are priced more highly by the market than low-volatility scenarios. Specifically, we conclude that the change of measure λ must be positively correlated with $W^{\mathbb{Q}}$.

3. The Rough Bergomi (rBergomi) model

Despite the inconsistency with VIX smiles pointed out above, let us nevertheless consider the simplest change of measure

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

assuming for simplicity, resp. as a first approximation, that $\lambda(s)$ is a deterministic function of s . Then, from (2.8), we would have

$$\begin{aligned}
 v_u &= \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{Q}}(u) \right) \\
 &\quad \times \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{1}{(u-s)^\gamma} \lambda(s) ds \right\} \\
 &= \xi_t(u) \mathcal{E} \left(\eta \tilde{W}_t^{\mathbb{Q}}(u) \right) \quad (3.1)
 \end{aligned}$$

where by definition, $\xi_t(u) = \mathbb{E}^{\mathbb{Q}} [v_u | \mathcal{F}_t]$. Moreover, the forward variance curve

$$\xi_t(u) = \mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t] \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{1}{(u-s)^\gamma} \lambda(s) ds \right\}$$

is the product of two terms: $\mathbb{E}^{\mathbb{P}} [v_u | \mathcal{F}_t]$ which depends on the history of the driving Brownian motion as explained earlier, and a term which depends on the price of risk $\lambda(s)$.

The model (3.1) is a non-Markovian generalization of the Bergomi model (1.5) which we might dub a *rough Bergomi* (or *rBergomi*) model. Specifically, this rBergomi model is non-Markovian in the instantaneous variance v_t : $\mathbb{E}^{\mathbb{Q}} [v_u | \mathcal{F}_t] \neq \mathbb{E}^{\mathbb{Q}} [v_u | v_t]$ but is Markovian in the (infinite-dimensional) state vector $\mathbb{E}^{\mathbb{Q}} [v_u | \mathcal{F}_t] = \xi_t(u)$.

Note also that with (3.1), we have achieved the aim we set out in the introduction by replacing the exponential kernels in the Bergomi model (1.5) with a power-law kernel. We may therefore expect that the rBergomi model will generate a realistic term structure of ATM volatility skew.

The observed anticorrelation between price moves and volatility moves may be modelled naturally, just as in the conventional Bergomi model, 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.

Remark 3.1 Note that in the rBergomi model, option prices (and in fact the statistics of the process) as seen from time t depend only on the forward variance curve $\xi_t(u)$ and the three parameters H , η and ρ . In this sense there is no explicit dependence on t which is essentially just a label.

3.1. Re-interpretation of the conventional Bergomi model

According to Bergomi (2005), the conventional Bergomi model is a *market model*, by which it is meant that $\xi_t(u)$ can be any given initial forward variance swap curve consistent with market prices. However, for the Bergomi model to properly describe the evolution of this curve, $\xi_t(u) = \mathbb{E}[v_u | \mathcal{F}_t]$ should be consistent with the assumed dynamics; in this sense, a conventional n -factor Bergomi model is not self-consistent in general.

Viewed from the perspective of the rBergomi 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 (1.5) 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 rBergomi model.

4. Simulation of the rBergomi model

To simplify notation, we set the origin of the simulation to be $t = 0$ and drop the explicit reference to the pricing measure \mathbb{Q} . From (3.1), the model to be simulated is

$$S_t = S_0 \mathcal{E} \left(\int_0^t \sqrt{v_u} dZ_u \right) \\ v_u = \xi_0(u) \mathcal{E} \left(\eta \sqrt{2H} \int_0^u \frac{1}{(u-s)^\gamma} dW_s \right) = \xi_0(u) \mathcal{E} \left(\eta \tilde{W}_u \right).$$

where \tilde{W} is a Volterra process[†] with the scaling property $\text{Var}[\tilde{W}_u] = u^{2H}$. So far \tilde{W} behaves just like fBm. However, the dependence structure is different. Specifically, for $v > u$,

$$\mathbb{E}[\tilde{W}_v \tilde{W}_u] = u^{2H} G\left(\frac{v}{u}\right)$$

[†]This is identical up to a constant factor to the definition of Comte and Renault (1996).

where, for $x \geq 1$ and with $\gamma = 1/2 - H$,

$$G(x) = 2H \int_0^1 \frac{ds}{(1-s)^\gamma (x-s)^\gamma} \\ = \frac{1-2\gamma}{1-\gamma} x^\gamma {}_2F_1(1, \gamma, 2-\gamma, x) \quad (4.1)$$

where ${}_2F_1(\cdot)$ denotes the confluent hypergeometric function.

Remark 4.1 The dependence structure of the Volterra process \tilde{W} is markedly different from that of fBm with the Molchan–Golosov kernel $K(u, s)$ given by

$$K(u, s) = c_H \frac{1}{(u-s)^\gamma} {}_2F_1\left(1/2 - H, H - 1/2, H + 1/2, \frac{s-u}{s}\right), \\ 0 < s < t,$$

for some constant c_H . In particular, for small H , correlations drop precipitously as the ratio u/v moves away from 1.

We also need covariances of the Brownian motion Z with the Volterra process \tilde{W} . With $v \geq u$, these are given by

$$\mathbb{E}[\tilde{W}_v Z_u] = \rho D_H \left\{ v^{H+1/2} - (v-u)^{H+1/2} \right\}$$

and

$$\mathbb{E}[Z_v \tilde{W}_u] = \rho D_H u^{H+1/2}$$

where for future convenience, we have defined the constant,

$$D_H = \frac{\sqrt{2H}}{H+1/2}.$$

These two formulae may be conveniently combined as

$$\mathbb{E}[\tilde{W}_v Z_u] = \rho D_H \left\{ v^{H+1/2} - (v - \min(u, v))^{H+1/2} \right\}.$$

Lastly, of course, for $v \geq u$, $\mathbb{E}[Z_v Z_u] = u$.

With m the number of time steps and n the number of simulations, our rBergomi model simulation algorithm may then be summarized 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 $m \times 2n$ 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.

The simulation procedure we have described is unsurprisingly very slow because of the high number of matrix-vector multiplications with a lower triangular but otherwise dense matrix. We leave the search for faster simulation techniques based on the specific structure of the problem, including the specific choice of the correlation structure between Z and \tilde{W} for future research.

5. Consistency of the rBergomi model with the SPX volatility surface

As mentioned earlier, our simulation of the rBergomi model is very slow. However, the model has only three parameters:

H , η and ρ . Although there is a little bit of art involved, these parameters may be identified relatively straightforwardly from the observed implied volatility surface by exploiting the following observations:

- H controls the decay of the term structure of the ATM volatility skew $\psi(\tau)$ for small τ .
- Fixing H , the product $\rho\eta$ is fixed by the level of $\psi(\tau)$ for larger τ .
- Focusing on a single expiration and fixing the product $\rho\eta$, the minimum of the smile moves lower and the position of the minimum moves to the right as ρ is made more negative.

rBergomi parameters can thus be guessed in practice.

We will now show that on two particular days in history, the rBergomi model was surprisingly consistent with the observed volatility surface. Fits for other days we tried are not always as impressive as these two but nevertheless visibly superior to fits of conventional Markovian stochastic volatility models.

5.1. Estimation of the variance swap curve

Variance swaps are actively traded so in principle, computation of the forward variance swap curve should be straightforward. In practice, however, it is not easy to obtain high-quality variance swap quote data and in any case, the bid/ask spread is wide. We thus choose to proxy the value of a τ -maturity variance swap by the value of a τ -expiration log contract as explained for example in Chapter 11 of Gatheral (2006). To price the log contract for a particular expiration τ requires us to know the prices of τ -expiration options for all strikes; of course prices are only quoted for a finite number of strikes. We therefore choose to interpolate and extrapolate observed implied volatilities using the arbitrage-free SVI parameterization of the volatility surface as explained in Gatheral and Jacquier (2014). For any given day, we obtain the closing prices of SPX options for all available strikes and expirations from OptionMetrics (www.optionmetrics.com) via Wharton Research Data Services (WRDS). Having estimated variance swaps to each expiration, we interpolate total variances using a monotonic spline to estimate variance swaps for intermediate dates. This allows us in turn to estimate the forward variance swap curve.

One subtlety is that by choosing SVI to interpolate and extrapolate, we may be assuming a smile that is inconsistent with the one generated by the rBergomi model, and therefore that the forward variance curve may not be accurate. The practical effect of this is that ATM implied volatilities are not matched in the first pass, with good agreement for very short expirations but rather less good agreement as time to expiry increases. A simple iteration on the forward variance curve soon reaches a fixed point that achieves consistency between model ATM volatilities and market ATM volatilities.

To summarize, we use the following procedure to compute parameters for the rBergomi model from market data.

- (1) Use SVI to interpolate and extrapolate market prices of put and call options at any strikes and expirations;
- (2) Compute the price of the log-strip contract (for any expiration) and extract the forward variance curve by differentiation;
- (3) Plug the forward variance curve and guessed values of H , ρ and η into the rBergomi model;
- (4) Compute option prices and ATM implied volatilities in the rBergomi model;
- (5) Compare with observed ATM implied volatilities and adjust the forward variance curve (and possibly the parameters H , ρ and η) and go back to (3).

We follow the procedure until we reach a fixed point in forward variance. In (5), we essentially adjust the forward variance curve by shifting and scaling, but this is a purely *ad hoc* choice.

5.2. Fits to two specific days in history

5.2.1. February 4, 2010. For our first comparison of the model to SPX options data, we choose February 4, 2010, a day when the ATM volatility term structure happened to be pretty flat. With guessed parameters $H = 0.07$, $\eta = 1.9$, $\rho = -0.9$, we obtain the impressive fit shown in figure 4. Only three parameters to get a very good fit to the whole SPX volatility surface, including the shortest dated smile (figure 5).

In figures 6 and 7, we see just how well the rBergomi model can match empirical skews and vols. Recall also that the parameters we used are just guesses!

5.2.2. August 14, 2013. For our second comparison, we choose a date just prior to an expiration date for SPX options. Rather than choosing the last Thursday of trading, we examine the volatility surface as of the close on the final Wednesday prior to expiration so that the shortest expiration smile is more meaningful; the latest such date available to us in our OptionMetrics data-set is Wednesday August 14, 2013. With guessed parameters $H = 0.05$, $\eta = 2.3$, $\rho = -0.9$, we obtain the fit shown in figure 8. Once again, only three parameters to get a very good fit to the whole SPX volatility surface, including the shortest dated smile (from options with only one day of trading left).

5.2.3. Jump-like behaviour of the rBergomi price process.

It has often been claimed that jumps are required to explain the observed extreme short-dated smile in SPX. In particular, in Carr and Wu (2003), Carr and Wu determine whether or not there are jumps in the asset process, and if so, whether such jumps are finite or infinite activity. They determine based on their analysis that jumps are required to generate the smiles observed in SPX. However, the class of processes that Carr and Wu consider is too restrictive, excluding models like rBergomi where the out-of-the-money smile explodes as time to expiration $\tau \rightarrow 0$. It is apparent from figures 4 and 8 that the rBergomi model (where the price process is continuous) generates smiles consistent with those observed empirically even for very short expirations; there is no need for jumps.

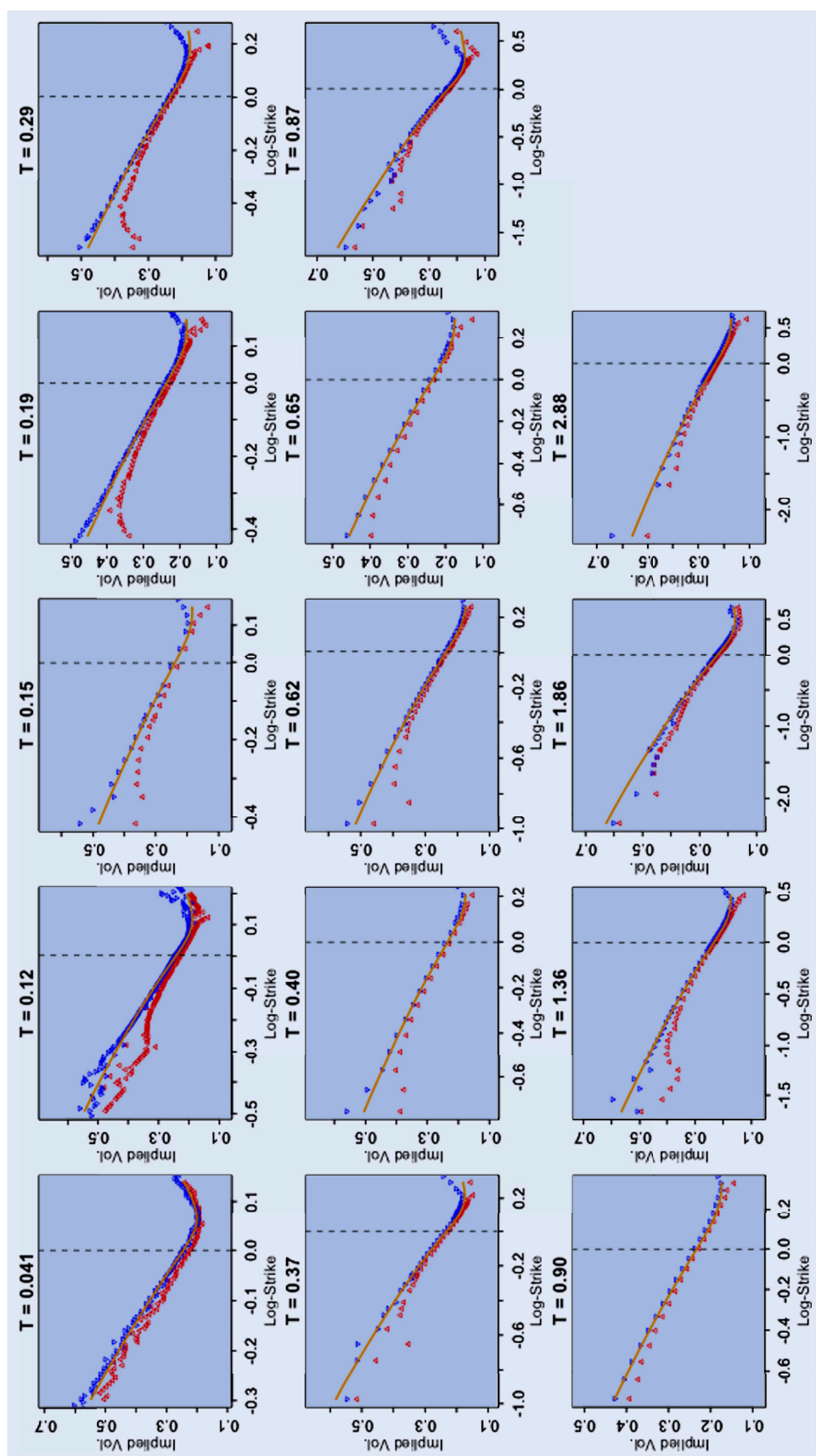


Figure 4. SPX smiles as of February 4, 2010: Red and blue points represent bid and offer SPX implied volatilities; orange smiles are from the rBergomi simulation.

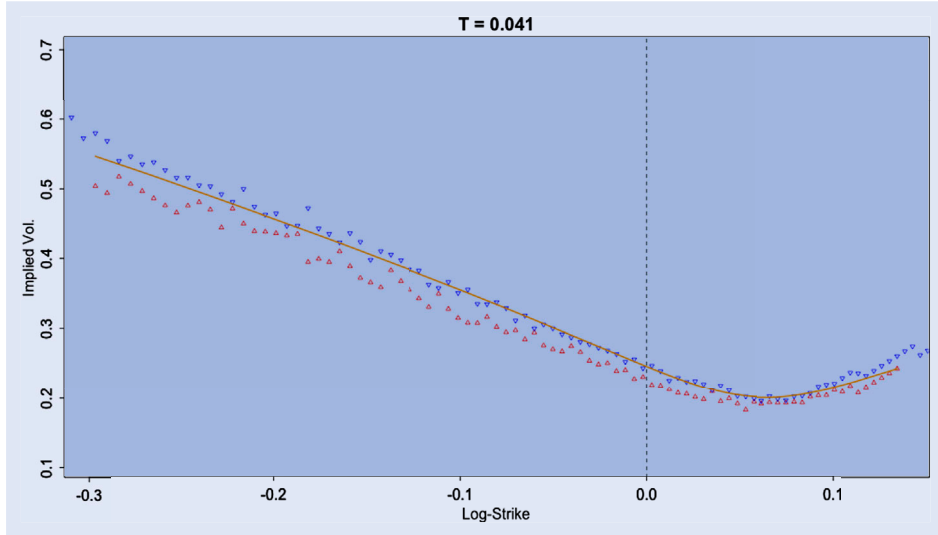


Figure 5. Shortest dated smile as of February 4, 2010: Red and blue points represent bid and offer SPX implied volatilities; orange smile is from the rBergomi simulation.

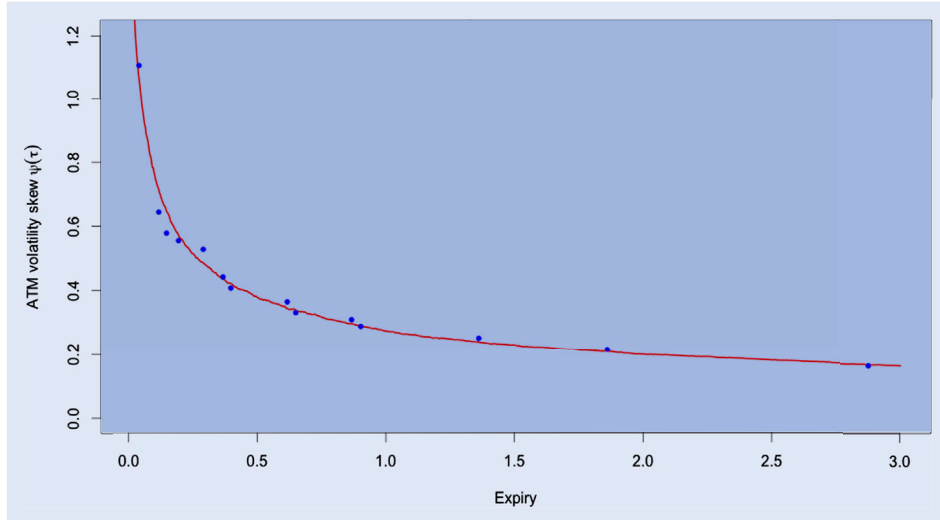


Figure 6. As of February 4, 2010: Blue points are empirical skews; the red line is from the rBergomi simulation. Time measured in years.

6. The rBergomi model and VIX options

We pointed out earlier in section 2.1 that observed VIX smiles are inconsistent with the rBergomi model. Nevertheless, even if the rBergomi model is misspecified, it may be possible to impute its parameters H and η by examining the term structure of VIX variance swaps[†]; if VIX corresponds to volatility, then VIX of VIX should correspond to ‘volatility of volatility’. If it

[†]The VIX log-strip forms the basis for the VVIX (VIX of VIX) index computation. Indeed, following CBOE (www.cboe.com), the VVIX term structure is computed every day (t) for various maturities (T) of VIX options using the usual log-strip formula that is used for the construction of VIX. More specifically, given $T > t$ and assuming that VIX options with expiry T are traded, the VVIX $_{t,T}$ is given by

$$\text{VVIX}_{t,T}^2(T-t) = -2\mathbb{E}_t \left[\log \sqrt{\zeta(T)} - \log \sqrt{\zeta(t)} \right],$$

where $\zeta(s)$ denotes the square of VIX at s and $\mathbb{E}_t \log \sqrt{\zeta(T)}$ can be expressed in terms of put and call prices on VIX with expiry T . The

were possible to impute H and η in this way, model calibration would obviously be considerably easier.

Denote the terminal value of the VIX futures by $\sqrt{\zeta(T)}$. Then, by definition,[‡]

$$\zeta(T) = \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}[v_u | \mathcal{F}_T] du.$$

where Δ is one month. In the rBergomi model,

$$v_u = \xi_t(u) \mathcal{E} \left(\eta \sqrt{2H} \int_t^u \frac{dW_s}{(u-s)^\gamma} \right)$$

with $\gamma = 1/2 - H$. Instantaneous variances v_u are thus lognormally distributed. It should therefore be a good approximation (and so it turns out) to assume that the VIX payoff and its square $\zeta(T)$ are also lognormally distributed. In that case, the terminal

usual VVIX index (at a given t) then corresponds to VVIX $_{t,t+\Delta}$ for Δ equal to one month.

[‡]See Chapter 11 of Gatheral (2006) for more details.

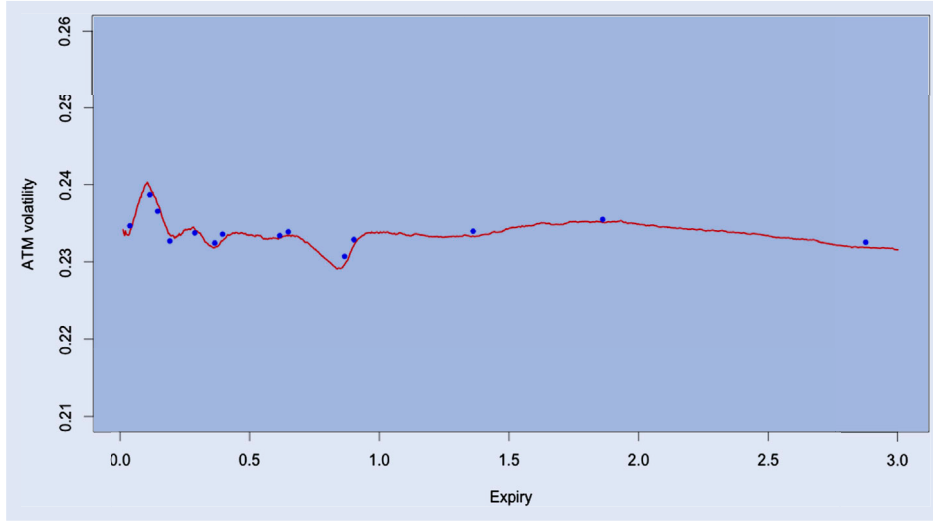


Figure 7. As of February 4, 2010: Blue points are empirical ATM volatilities; the red line is from the rBergomi simulation. Time measured in years.

distribution of $\zeta(T)$ is completely determined by $\mathbb{E}[\zeta(T)|\mathcal{F}_t]$ and $\text{Var}[\log \zeta(T)|\mathcal{F}_t]$.

It is immediate that

$$\mathbb{E}[\zeta(T)|\mathcal{F}_t] = \frac{1}{\Delta} \int_T^{T+\Delta} \xi_t(u) du.$$

To estimate the conditional variance of $\zeta(T)$, we approximate the arithmetic mean by the geometric mean as follows:

$$\zeta(T) \approx \exp \left\{ \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}[\log v_u | \mathcal{F}_T] du \right\}.$$

After some computation detailed in appendix 3, we obtain

$$\text{Var}[\log \zeta(T)|\mathcal{F}_t] \approx \eta^2 (T-t)^{2H} f^H \left(\frac{\Delta}{T-t} \right)$$

where

$$f^H(\theta) = \frac{D_H^2}{\theta^2} \int_0^1 \left[(1+\theta-x)^{1/2+H} - (1-x)^{1/2+H} \right]^2 dx. \quad (6.1)$$

It is straightforward to show that $f^H(\theta) \rightarrow 1$ as $\theta \rightarrow 0$ which is the limit in which $\zeta(T) \rightarrow v_T$. In appendix 3, we show further how to express $f^H(\theta)$ explicitly in terms of the hypergeometric function. However, the above form (6.1) is more convenient for computation.

The VIX variance swaps (VVIX²) are then given by

$$\begin{aligned} \text{VVIX}_{t,T}^2 (T-t) &\approx \text{Var} \left[\log \sqrt{\zeta(T)} \middle| \mathcal{F}_t \right] \\ &\approx \frac{1}{4} \eta^2 (T-t)^{2H} f^H \left(\frac{\Delta}{T-t} \right). \end{aligned} \quad (6.2)$$

VIX variance swaps may also be estimated directly from market prices of options on VIX using the log-strip in the usual way as

$$\text{VVIX}_{t,T}^2 (T-t) = -2 \mathbb{E} \left[\log \sqrt{\zeta(T)} - \log \sqrt{\zeta(t)} \middle| \mathcal{F}_t \right]. \quad (6.3)$$

By comparing the model VVIX term structure (6.2) with the market VVIX term structure (6.3), we can in principle fix the model parameters H and η .

6.1. The VVIX term structure in practice

Recall that the parameters we used to obtain the reasonably impressive SPX fits of section 5.2.1 were just guessed; specifically these parameters were as follows:

	H	η
February 4, 2010	0.07	1.9
August 14, 2013	0.05	2.3

In figure 9, we show plots of equation (6.2) with the above parameters, superimposed on the empirically estimated variances of VIX. At least for February 4, 2010, given the qualitative agreement between the shape of the curve (6.2) and empirical estimates, it might be possible to argue some consistency of the rBergomi model with observation; for August 14, 2013 however, there is not even qualitative agreement. Whether this disagreement between model and market should be ascribed to a misspecified model, wrong market prices, or indeed both of these, is left for future research.

More specifically, assuming the RFSV model to be a good description of the instantaneous variance process under \mathbb{P} , the aim of the present paper is to explore the consequences for option pricing under \mathbb{Q} . The rBergomi model results from assuming a deterministic change of measure. Figure 9 shows that not only can the rBergomi model not generate the observed upward-sloping VIX option smile, as explained in section 2.1, it cannot explain the level of VIX volatility either. This could be either because the RFSV model is misspecified, failing to capture important elements of the dynamics of instantaneous variance, or because the market prices of options on VIX are inconsistent with the market prices of options on SPX.

7. Volatility of volatility expansion

As simulation of the rBergomi model is so slow, one potential alternative is to estimate rBergomi parameters using an asymptotic expansion. Given a model expressed in forward variance

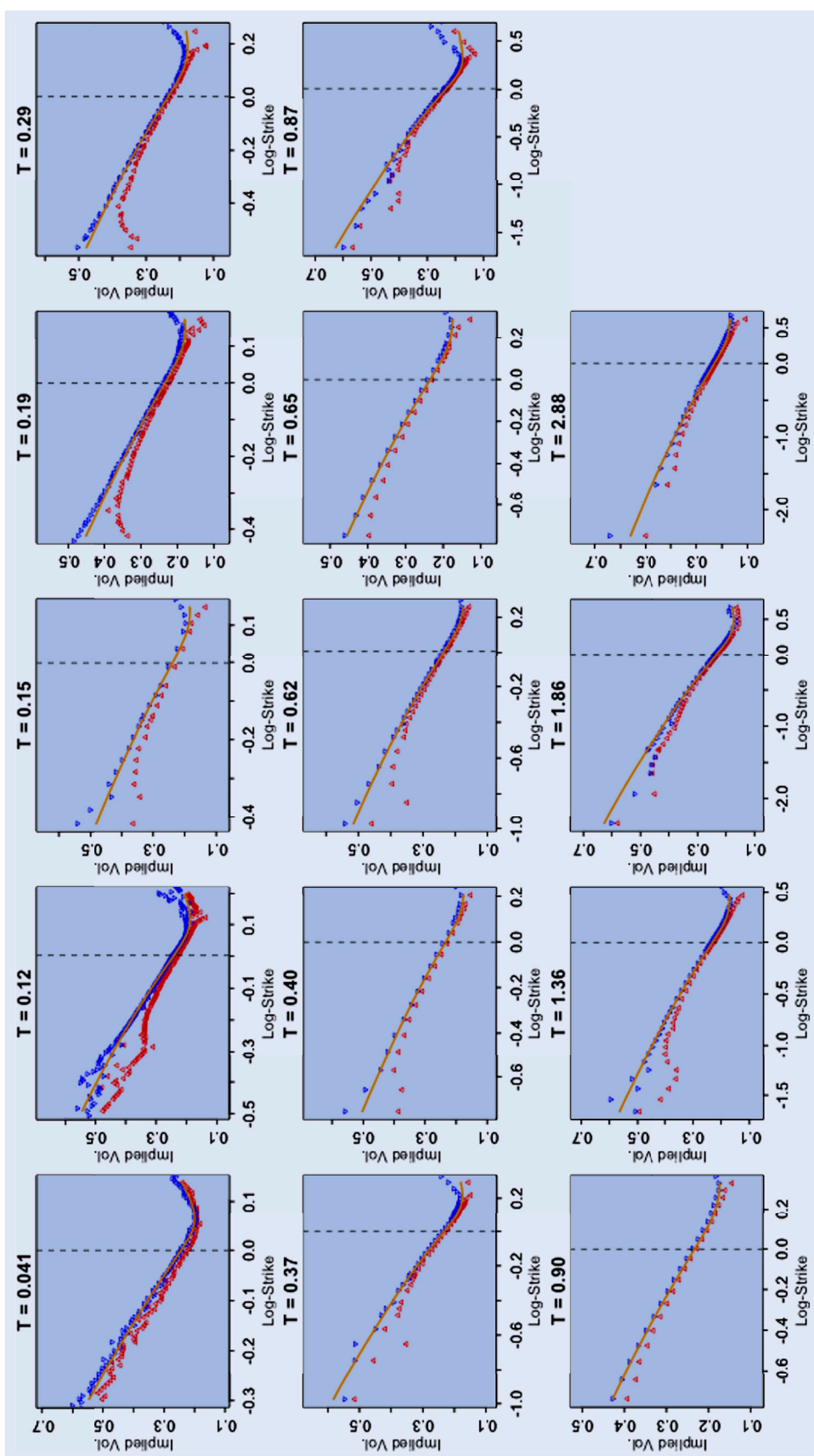


Figure 8. SPX smiles as of August 14, 2013: Red and blue points represent bid and offer SPX implied volatilities; orange smiles are from the rBergomi simulation.

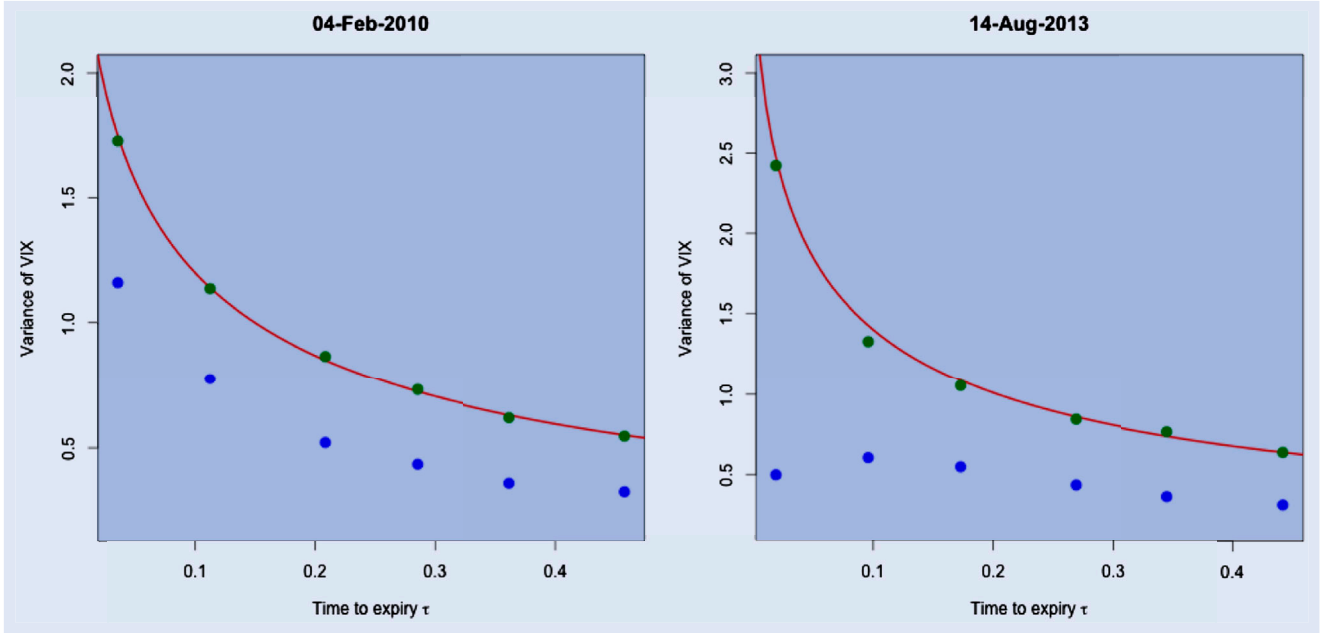


Figure 9. Empirical VVIX term structure data (blue points) and rBergomi (using (6.2), red line and from simulation, green points) estimates of $\text{Var}(\text{VIX})$ as of February 4, 2010 (left) and August 14, 2013 (right). Time measured in years.

curve form, Bergomi and Guyon (2012) derive the following expression for the Black–Scholes implied volatility smile to second order in volatility of volatility:

$$\sigma_{\text{BS}}(k, t) = \hat{\sigma}_T + S_T k + C_T k^2 + O(\eta^3) \quad (7.1)$$

where

$$\begin{aligned} \hat{\sigma}_T &= \sqrt{\frac{w}{T}} \left\{ 1 + \frac{1}{4w} C^{x\xi} + \frac{1}{32w^3} \right. \\ &\quad \times \left. \left(12(C^{x\xi})^2 - w(w+4)C^{\xi\xi} + 4w(w-4)C^\mu \right) \right\} \\ S_T &= \sqrt{\frac{w}{T}} \left\{ \frac{1}{2w^2} C^{x\xi} + \frac{1}{8w^3} \left(4wC^\mu - 3(C^{x\xi})^2 \right) \right\} \\ C_T &= \sqrt{\frac{w}{T}} \frac{1}{8w^4} \left(4wC^\mu + wC^{\xi\xi} - 6(C^{x\xi})^2 \right) \end{aligned} \quad (7.2)$$

and

$$w = \int_0^T \xi_0(s) ds$$

is total variance to expiration T . The autocorrelation functionals $C^{x\xi}$, $C^{\xi\xi}$ and C^μ have the following explicit expressions:

$$\begin{aligned} C^{x\xi} &= \int_0^T dt \int_t^T du \frac{\mathbb{E}[dx_t d\xi_t(u)]}{dt} \\ C^{\xi\xi} &= \int_0^T dt \int_t^T ds \int_t^T du \frac{\mathbb{E}[d\xi_t(s) d\xi_t(u)]}{ds} \\ C^\mu &= \int_0^T dt \int_t^T du \frac{\mathbb{E}[dx_t d\xi_t(u)]}{dt} \frac{\delta C_t^{x\xi}}{\delta \xi_t(u)} \end{aligned} \quad (7.3)$$

where the notation $\delta/\delta \xi_t(u)$ denotes a functional derivative.

In the case of the rBergomi model (3.1), we have

$$\begin{aligned} \frac{dS_t}{S_t} &= \sqrt{\xi_t(t)} dZ_t \\ \frac{d\xi_t(u)}{\xi_t(u)} &= \eta \sqrt{2H} \frac{dW_t}{(u-t)^\gamma} \end{aligned}$$

with $\mathbb{E}[dZ_t dW_t] = \rho dt$ so that

$$\frac{\mathbb{E}[dx_t d\xi_t(u)]}{dt} = \rho \eta \sqrt{2H} \sqrt{\xi_t(t)} \frac{\xi_t(u)}{(u-t)^\gamma}.$$

The various autocorrelation functionals (7.3) may then be computed; explicit computations are presented in appendix 2.

7.1. Special case: flat variance curve

The special case $\xi_0(u) = \bar{\sigma}^2$, $u \geq 0$ where the initial forward variance curve is flat, is particularly instructive. First, from (B1), we have

$$\begin{aligned} C^{x\xi} &= \rho \eta \sqrt{2H} \int_0^T \sqrt{\xi_t(s)} ds \int_s^T \xi_t(u) \frac{du}{(u-s)^\gamma} \\ &= \rho \eta \bar{\sigma}^3 E_H T^{H+3/2}. \end{aligned} \quad (7.4)$$

where we have further defined

$$E_H = \frac{D_H}{H+3/2}.$$

Also, $w = \bar{\sigma}^2 T$. Substituting back into (7.2) and then (7.1) gives, to first order in η ,

$$\sigma_{\text{BS}}(k, t) = \bar{\sigma} + \frac{\rho \eta}{2} E_H \frac{1}{T^{1/2-H}} \left(k + \frac{w}{2} \right) + O(\eta^2) \quad (7.5)$$

In particular, we see that to first order in η , the term structure of ATM volatility skew is given by

$$\psi(\tau) = \left. \frac{\partial \sigma_{\text{BS}}^2(k, \tau)}{\partial k} \right|_{k=0} = \frac{\rho \eta}{2} E_H \frac{1}{\tau^\gamma}$$

with $\gamma = 1/2 - H$. Similarly, substituting $\xi_0(u) = \bar{\sigma}^2$ into (B2) and (B3), respectively, gives the terms required for computation of the second-order contribution:

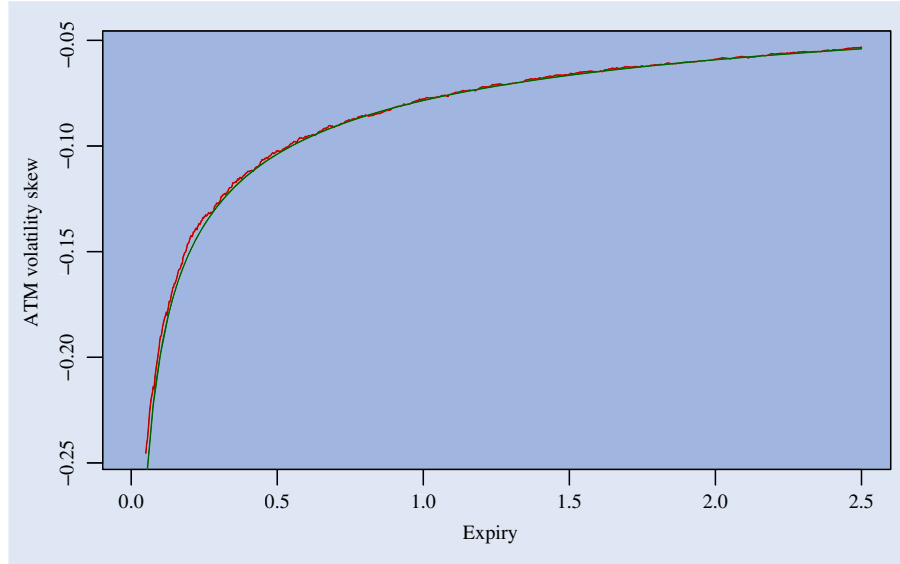


Figure 10. The Bergomi–Guyon second order ATM skew approximation is in green; ATM skews from Monte Carlo simulation are in red. Parameters used were $H = 0.1$, $\eta = 0.4$, $\rho = -0.85$, $\bar{\sigma} = 0.235$.

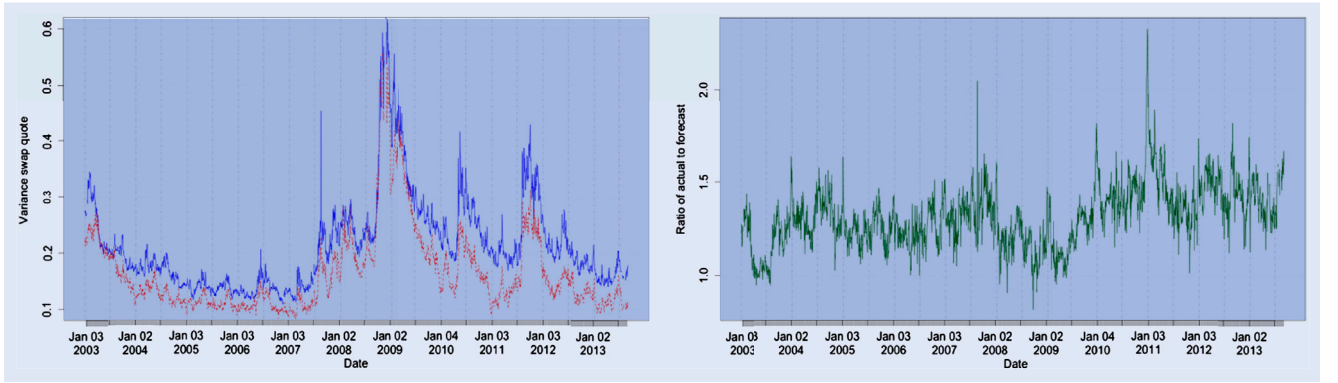


Figure 11. Plot (a) shows actual (proxy) 3-month variance swap quotes in blue vs forecast in red. Plot (b) shows the ratio between 3-month actual variance swap quotes and 3-month forecasts.

$$C^{\xi\xi} = \eta^2 \bar{\sigma}^4 D_H^2 \frac{T^{2+2H}}{2+2H}$$

$$C^\mu = \frac{1}{2} \rho^2 \eta^2 \bar{\sigma}^4 D_H^2 \left\{ 1 + \frac{\Gamma(H+3/2)^2}{\Gamma(2H+3)} \right\} \frac{T^{2+2H}}{2+2H}.$$

It follows that to second order in η , the term structure of ATM volatility skew is given by

$$\psi(\tau) = \frac{\rho \eta}{2} E_H \frac{1}{\tau^\gamma} + \frac{1}{4} \rho^2 \eta^2 \bar{\sigma} \tau^{2H} \times \left[\frac{D_H^2}{1+H} \left\{ 1 + \frac{\Gamma(H+3/2)^2}{\Gamma(2H+3)} \right\} - \frac{3}{2} E_H^2 \right]. \quad (7.6)$$

7.1.1. Numerical test. The dimensionless Bergomi–Guyon expansion parameter is $\lambda = \eta T^H$. When H is very small, $\lambda \sim \eta$ for all reasonable expirations; with $H < 0.1$ as in section 5.2.1, $\lambda \sim 1.9$ which is not small enough for the asymptotic expansion to converge, even ATM. With the much smaller value $\eta = 0.4$, we see in figure 10 very good agreement between the Bergomi–Guyon asymptotic skew formula (7.6) and the simulation.

We thus conclude that both our Bergomi–Guyon computations and the simulation are likely to be correct. Sadly however, the Bergomi–Guyon expansion does not converge with values of η consistent with the SPX volatility surface, so the Bergomi–Guyon expansion is not useful in practice for calibration of the rBergomi model.

8. Forecasting the variance swap curve

Recall that in the RFSV model (1.1), $\log v_t \approx 2 \nu W_t^H + C$ for some constant C . In Nuzman and Poor (2000), it is shown, assuming $H \in (0, 1/2)$, $\Delta > 0$, that $W_{t+\Delta}^H$ is conditionally Gaussian with conditional expectation[†]

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

[†]Trivially $\mathbb{E}[W_{t+\Delta}^H | \mathcal{F}_t] = W_t^H$ when either $\Delta = 0$ or $H = 1/2$. This corresponds to the singular behaviour of the integrand, as either $\Delta \rightarrow 0$ or $H \rightarrow 1/2$, with the necessary compensation given by $\Delta^{H+1/2} \cos(H\pi) \sim 0$ in these regimes.

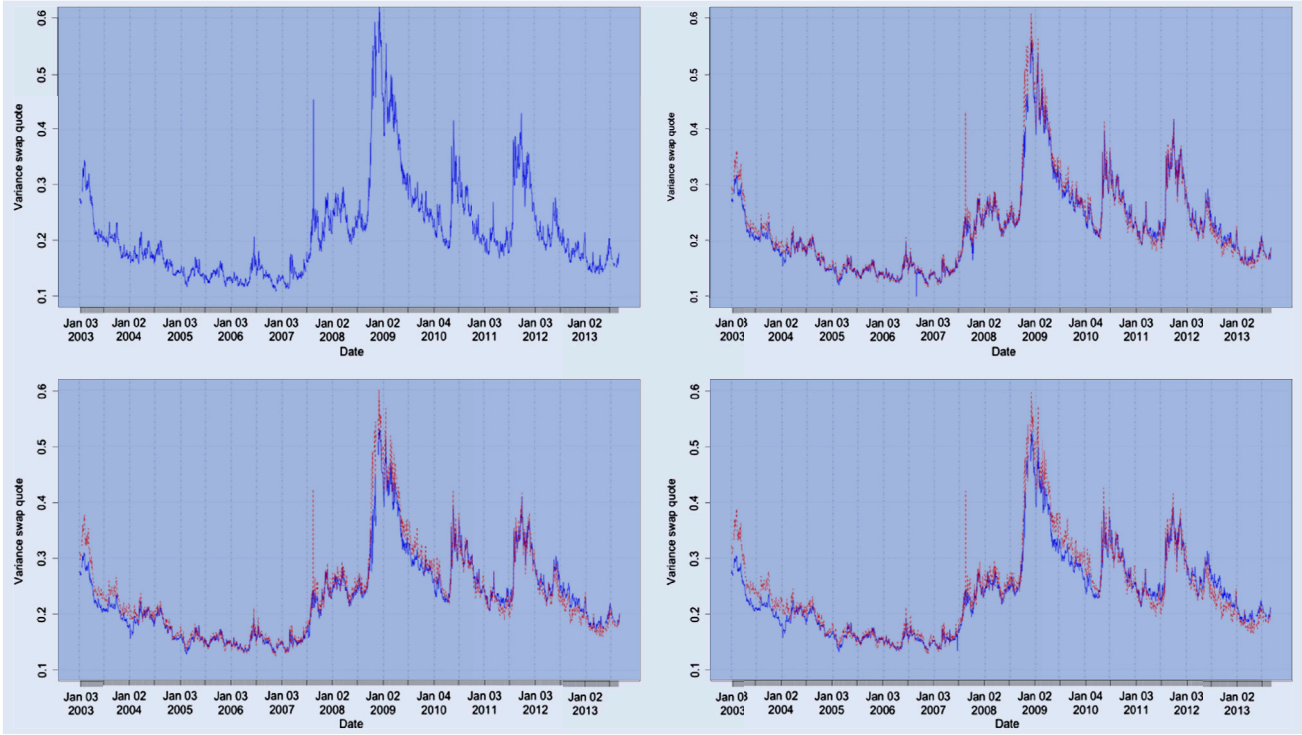


Figure 12. Plot (a) shows actual (proxy) 3-month variance swap quotes. The other 3 figures show actual variance swap quotes for 6, 9 and 12 month, respectively, in blue with forecast variance swap quotes multiplied by the 3-month actual to forecast ratio in red.

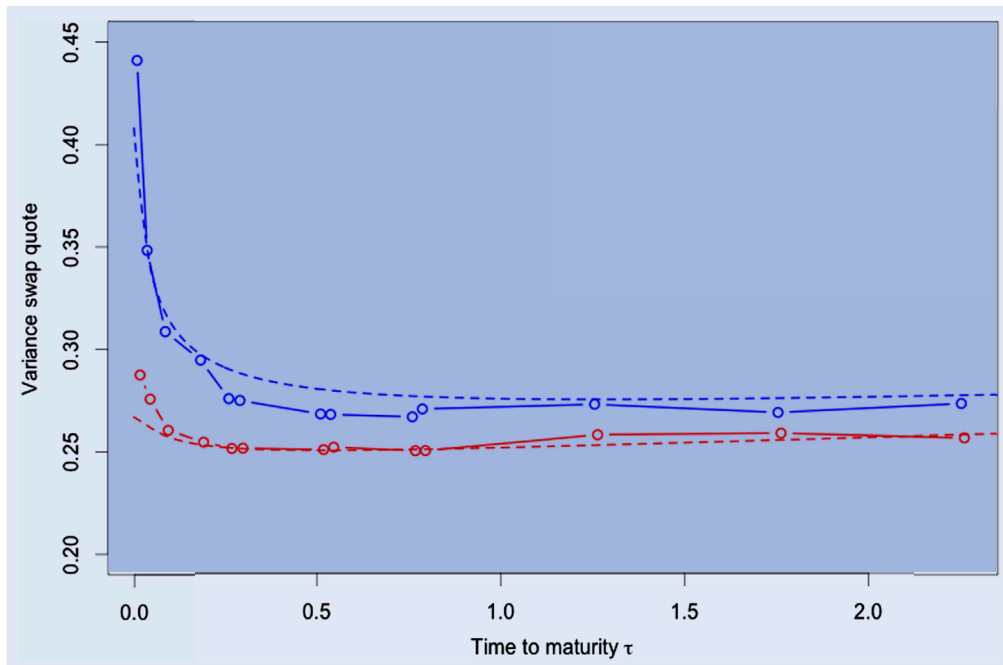


Figure 13. S&P variance swap curves as of September 12, 2008 (red) and September 15, 2008 (blue). The dashed curves are RFSV model forecasts rescaled by the 3-month ratio (1.29) as of the Friday close. Time measured in years.

and conditional variance

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

where

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

Thus, we obtain the following natural form for the RFSV predictor of the variance:

$$\mathbb{E}^{\mathbb{P}}[v_{t+\Delta} | \mathcal{F}_t] = \exp \left\{ \mathbb{E}^{\mathbb{P}}[\log(v_{t+\Delta}) | \mathcal{F}_t] + 2c v^2 \Delta^{2H} \right\} \quad (8.1)$$

where

$$\mathbb{E}^{\mathbb{P}}[\log v_{t+\Delta} | \mathcal{F}_t] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \times \int_{-\infty}^t \frac{\log v_s}{(t-s+\Delta)(t-s)^{H+1/2}} ds. \quad (8.2)$$

The fair value of a τ -maturity variance swap is given (approximately) by

$$\mathcal{V}_t(\tau) = \frac{1}{\tau} \int_t^{t+\tau} \mathbb{E}^{\mathbb{Q}}[v_s | \mathcal{F}_t] ds$$

where \mathbb{Q} is the risk neutral measure. If it were possible to ignore the change of measure so that

$$\mathbb{E}^{\mathbb{Q}}[v_s | \mathcal{F}_t] = \mathbb{E}^{\mathbb{P}}[v_s | \mathcal{F}_t],$$

it would be possible to forecast variance swap curves using (8.1). In fact, we will see that from the data, \mathbb{Q} is close to \mathbb{P} in this sense. We now proceed to compare forecast and actual variance swaps curves.

SPX variance curve forecasts are formed using the predictor (8.1) from the time series of daily realized variance estimates from same Oxford-Man data-set that was used in Gatheral *et al.* (2014).

As for market variance swap curves, although there is an active market, it is not easy to obtain high-quality variance swap quote data and in any case, the bid/ask spread is wide. We thus choose to proxy the value of a τ -maturity variance swap by the value of a τ -expiration log contract as explained for example in Chapter 11 of Gatheral (2006). To price the log contract for a particular expiration τ requires us to know the prices of τ -expiration options for all strikes; of course prices are only quoted for a finite number of strikes. We therefore choose to interpolate and extrapolate observed implied volatilities using the arbitrage-free SVI parameterization of the volatility surface as explained in Gatheral and Jacquier (2014). With closing prices of SPX options for all available strikes and expirations sourced from OptionMetrics (www.optionmetrics.com) via WRDS, we follow the procedure just described to compute proxy variance swap curves each day from January 4, 1996 to August 30, 2013, a total of 4443 days. We also need a sufficient history to be able to compute a forecast. We end up with 2681 days of forecast and actual variance swap curves from January 3, 2003 to August 31, 2013.

Plotting actual vs. forecast 3-month variance swap curves in figure 11, we immediately see that the actual variance swap curve is a factor (of roughly 1.4) higher than the forecast one, which we may attribute to overnight movements of the index. Recall that realized variance estimates are intraday from open to close. Integrated variance forecasts must therefore be rescaled to obtain close-to-close integrated variance forecasts as explained for example in Corsi *et al.* (2013) or alternatively using an econometric model such as the HEAVY model of Noreldin *et al.* (2012).

In figure 12, we see that the 6-month, 9-month and 12-month forecasts rescaled by the 3-month ratio of actual to forecast seem to be very consistent with actual variance swap quotes. This implies that although we can only forecast variance swap curves up to a factor, we can accurately forecast their shapes. We now demonstrate this further with two dramatic examples where the variance swap curve moved significantly from one

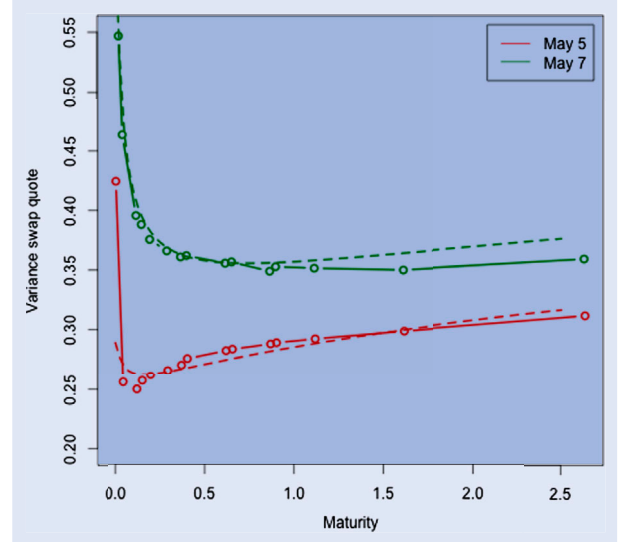


Figure 14. S&P variance swap curves as of May 5, 2010 (red) and May 7, 2010 (green). The dashed curves are RFSV model forecasts rescaled by the 3-month ratio (2.52) as of the close on Wednesday May 5. Time measured in years.

day to the next. We will see that in both these cases, the evolution of the variance swap curve seems to be consistent with our model paradigm.

8.1. The collapse of Lehman Brothers

As an interesting experiment,[†] consider the evolution of the S&P variance swap curve over the weekend of the collapse of Lehman Brothers. As of the market close on Friday September 12, 2008, it was generally expected that Lehman Brothers would be rescued over the weekend. As of the market close on Monday September 15 however, there had been no rescue and the market was in crisis. In figure 13, we plot the actual variance swaps curves as of the Friday and Monday market closes together with forecast curves rescaled by the 3-month ratio as of the close on Friday September 12 (which was 1.29). Perhaps surprisingly, it appears that most of the evolution of the variance swap curve may be explained by a single extra data point – intraday realized variance from the open to the close of trading on Monday September 15, 2008.

8.2. The Flash Crash

In the so-called Flash Crash of Thursday May 6, 2010, major US equity indices suddenly dropped by about 10% intraday only to recover within 30 min or so. Consequently, intraday realized variance for May 6 was much higher than normal. In figure 14, using the same methodology as in section 8.1, 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

[†]Suggested by Peter Leoni of KU Leuven.

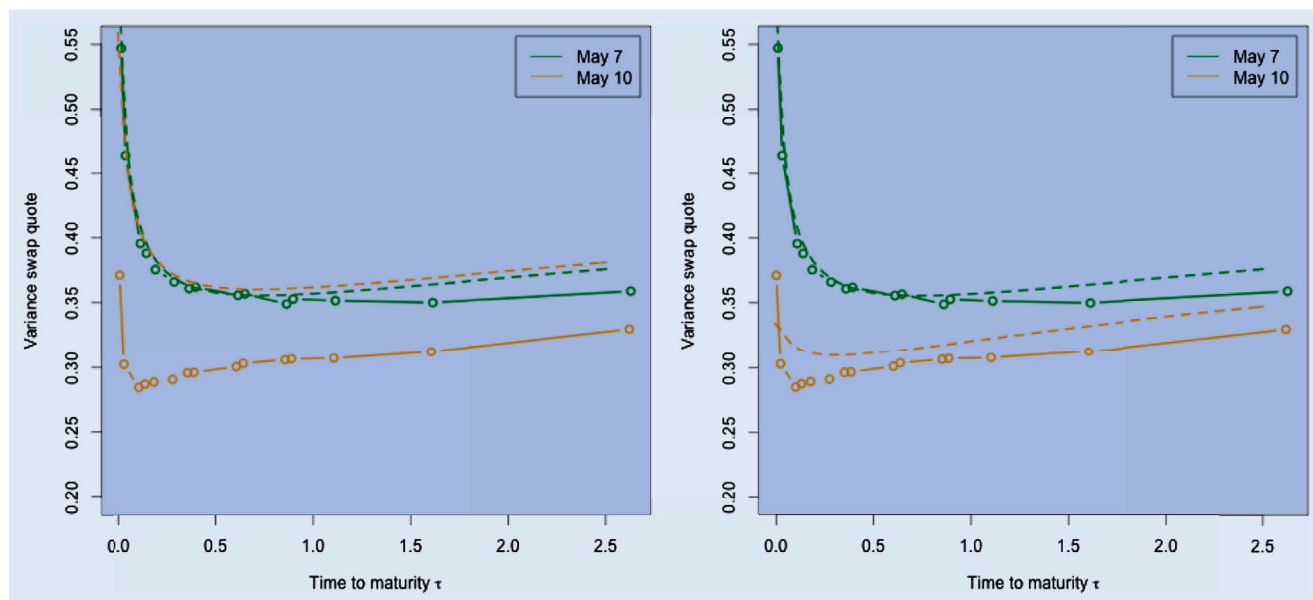


Figure 15. May 7 variance curves are in green; May 10 variance curves are in orange. Solid lines are actual curves and dashed lines are forecast curves. In plot (a), we see that the May 10 actual curve is completely inconsistent with a forecast based on a realized variance dataset that includes the Flash Crash. In contrast, in plot (b), we see that the Monday May 10 actual curve is consistent with a forecast using a data-set that excludes the Flash Crash.

Thursday May 6. In figure 15, we see that actual variance swap curve as of the following Monday close is no longer consistent with the forecast. However, if we drop the May 6 datapoint, we get a forecast that is much closer to the actual variance swap curve. The obvious explanation is that volatility traders realized over the weekend that the anomalous intraday price action of the Flash Crash should not influence future realized variance projections, adjusting index option quotes accordingly.

9. Summary and conclusions

The RFSV model of Gatheral *et al.* (2014) is remarkably consistent with the time series of realized variance for a wide range of different underlying assets. In this paper, we have shown that this model written under the physical measure \mathbb{P} leads naturally to an options pricing model under \mathbb{Q} that is remarkably consistent with the observed shape of the implied volatility surface in the particular case of SPX. A special case of this model where we assume a deterministic change of measure between \mathbb{P} and \mathbb{Q} turns out to be a non-Markovian extension of the well-known Bergomi model, which we consequently dub the Rough Bergomi (or rBergomi) model. The rBergomi model is particularly tractable and seems to fit the SPX volatility surface very well, despite our lack at this stage of an efficient computational algorithm. We computed terms Bergomi–Guyon expansion up to second order in volatility of volatility but the expansion parameter $\lambda = \eta \tau^H \approx 2$ required to fit SPX option prices is too big for this asymptotic expansion to be valid. However, we do not see agreement between the rBergomi model and the term structure of VIX volatility (VVIX).

Taken together, the present work and the econometric analysis of Gatheral *et al.* (2014) offer a (perhaps even the first) promising paradigm for the understanding of asset price forma-

tion all the way from a basic microstructure description at the order book level to option pricing. Not least, our framework allows for accurate prediction of the volatility surface from high-frequency price data. More efficient computational methods and a more thorough investigation of the market implied change of measure $d\mathbb{Q}/d\mathbb{P}$ are left for further research.

Acknowledgements

We are very grateful to the anonymous referees for their many detailed comments and suggestions which helped improve our paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

C.B. and P.F. gratefully acknowledge financial support by Deutsche Forschungsgemeinschaft through Research Grant BA 5484/1 and FR 2943/2.

References

- Alòs, E., León, J.A. and Vives, J., On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility. *Finance Stoch.*, 2007, **11**(4), 571–589.
- Bergomi, L., Smile dynamics II. *Risk*, October 2005, pp. 67–73.
- Bergomi, L. and Guyon, J., Stochastic volatility's orderly smiles. *Risk*, May 2012, pp. 60–66.
- Carr, P. and Wu, L., What type of process underlies options? A simple robust test. *J. Finance*, 2003, **58**(6), 2581–2610.
- Cheridito, P., Kawaguchi, H. and Maejima, M., Fractional Ornstein–Uhlenbeck processes. *Electron. J. Probab.*, 2003, **8**(3), 1–14.

- Comte, F., Coutin, L. and Renault, E., Affine fractional stochastic volatility models. *Ann. Finance*, 2012, **8**(2–3), 337–378.
- Comte, F. and Renault, E., Long memory continuous time models. *J. Econom.*, 1996, **73**(1), 101–149.
- Corcuera, J.M., Hedevarang, E., Pakkanen, M.S. and Podolskij, M., Asymptotic theory for Brownian semi-stationary processes with application to turbulence. *Stoch. Process. Appl.*, 2013, **123**(7), 2552–2574.
- Corsi, F., Fusari, N. and La Vecchia, D., Realizing smiles: Options pricing with realized volatility. *J. Financ. Econ.*, 2013, **107**(2), 284–304.
- Fukasawa, M., Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance Stoch.*, 2011, **15**(4), 635–654.
- Gatheral, J., *The volatility surface: A practitioner's guide*, 2006 (Wiley: Hoboken, NJ).
- Gatheral, J. and Jacquier, A., Arbitrage-free SVI volatility surfaces. *Quant. Finance*, 2014, **14**(1), 59–71.
- Gatheral, J., Jaisson, T. and Rosenbaum, M., Volatility is rough, 2014. Available at SSRN 2509457.
- Gatheral, J. and Oomen, R.C., Zero-intelligence realized variance estimation. *Finance Stoch.*, 2010, **14**(2), 249–283.
- Jaisson, T. and Rosenbaum, M., Limit theorems for nearly unstable Hawkes processes. *Ann. Appl. Probab.*, 2015, **25**(2), 600–631.
- McAleer, M. and Medeiros, M., Realized volatility: A review. *Econom. Rev.*, 2008, **27**(1), 10–45.
- Noureddin, D., Shephard, N. and Sheppard, K., Multivariate high-frequency-based volatility (heavy) models. *J. Appl. Econom.*, 2012, **27**(6), 907–933.
- Nuzman, C.J. and Poor, V.H., Linear estimation of self-similar processes via Lamperti's transformation. *J. Appl. Probab.*, 2000, **37**(2), 429–452.

Appendix 1. Volatility, variance and VIX

Here, we briefly recall some different notions of variance and volatility that are used throughout the paper. For the sake of concreteness, we assume a general stochastic volatility model

$$\frac{dS_t}{S_t} = \mu_t dt + \sqrt{v_t} dZ_t$$

for some continuous stochastic *instantaneous variance* process v_t and a Brownian motion Z_t —all understood under the physical measure \mathbb{P} . *Integrated variance* corresponds to the quadratic variation of $\log S_t$ and thus is given by $\int_t^T v_s ds$ in the above model. *Realized variance* is the most familiar and widely used of many possible statistical estimators of integrated variance. (See for example McAleer and Medeiros (2008) and Gatheral and Oomen (2010) for more details.) In particular, the Oxford Man Institute provides a large database of daily estimates of integrated variance for a number of indices, including realized variance and realized kernel estimates.

Forward variance is the expectation under the pricing measure of future instantaneous variance. That is,

$$\xi_t(u) = \mathbb{E}^{\mathbb{Q}}[v_u | \mathcal{F}_t], \quad u \geq t.$$

The *variance swap* is a forward contract on realized variance, the fair strike of which is given by

$$\mathcal{V}_t(T) = \int_t^T \xi_t(u) du.$$

Volatility is the square root of variance so for example, instantaneous volatility is the square root of instantaneous variance. The CBOE Volatility Index (VIX) is given by a computation that approximates the square root of the fair strike of a one-month variance swap on the S&P index (for more details see www.cboe.com and for example Chapter 11 of Gatheral (2006)). Analogously, the CBOE VVIX index approximates the square root of the fair strike of a one-month variance swap on the VIX index. Finally, *implied volatility* is the value of the volatility parameter in the Black–Scholes formula required to match the market price of a European option.

Appendix 2. Computation of Bergomi–Guyon autocorrelation functionals in the rBergomi model

Computation of $C^{x\xi}$

$$\begin{aligned} C_t^{x\xi} &= \int_t^T ds \int_s^T du \frac{\mathbb{E}[dx_s d\xi_s(u)]}{ds} \\ &= \rho \eta \sqrt{2H} \int_t^T ds \int_s^T \mathbb{E}[\sqrt{\xi_s(s)} \xi_s(u) | \mathcal{F}_t] \frac{du}{(u-s)^\gamma} \\ &= \rho \eta \sqrt{2H} \int_t^T \sqrt{\xi_t(s)} ds \int_s^T \xi_t(u) \frac{du}{(u-s)^\gamma} + \mathcal{O}(\eta^3). \end{aligned} \quad (\text{B1})$$

Computation of $C^{\xi\xi}$ By definition,

$$\begin{aligned} C^{\xi\xi} &= \int_0^T dt \int_t^T du \int_t^T ds \frac{\mathbb{E}[d\xi_t(s) d\xi_t(u)]}{dt} \\ &= \int_0^T dt \int_t^T du \int_t^T ds \frac{\eta^2 2H}{(u-t)^\gamma (s-t)^\gamma} \xi_t(s) \xi_t(u) \\ &= \eta^2 2H \int_0^T dt \left(\int_t^T \frac{\xi_t(u)}{(u-t)^\gamma} du \right)^2 \\ &= \eta^2 2H \int_0^T dt \left(\int_t^T \frac{\xi_0(u)}{(u-t)^\gamma} du \right)^2 + \mathcal{O}(\eta^4). \end{aligned} \quad (\text{B2})$$

Computation of C^μ

By definition,

$$C^\mu = \int_0^T dt \int_t^T du \frac{\mathbb{E}[dx_t d\xi_t(u)]}{dt} \frac{\delta C_t^{x\xi}}{\delta \xi_t(u)}$$

and from (B1) above,

$$\begin{aligned} \frac{\delta C_t^{x\xi}}{\delta \xi_t(v)} &= \rho \eta \sqrt{2H} \left\{ \int_t^T ds \sqrt{\xi_t(s)} \frac{1}{(v-s)^\gamma} \mathbf{1}_{v>s} \right. \\ &\quad \left. + \frac{1}{2\sqrt{\xi_t(v)}} \int_v^T \xi_t(u) \frac{du}{(u-v)^\gamma} \right\} \\ &= \rho \eta \sqrt{2H} \left\{ \int_t^v ds \sqrt{\xi_t(s)} \frac{1}{(v-s)^\gamma} \right. \\ &\quad \left. + \frac{1}{2\sqrt{\xi_t(v)}} \int_v^T \xi_t(u) \frac{du}{(u-v)^\gamma} \right\}. \end{aligned}$$

Thus,

$$\begin{aligned} C^\mu &= \rho^2 \eta^2 2H \int_0^T \sqrt{\xi_t(t)} dt \int_t^T du \frac{\xi_t(u)}{(u-t)^\gamma} \\ &\quad \times \left\{ \int_t^u ds \sqrt{\xi_t(s)} \frac{1}{(u-s)^\gamma} \right. \\ &\quad \left. + \frac{1}{2\sqrt{\xi_t(u)}} \int_u^T \xi_t(s) \frac{ds}{(s-u)^\gamma} \right\} \\ &= \rho^2 \eta^2 2H \int_0^T \sqrt{\xi_t(t)} dt \int_t^T \frac{du}{(u-t)^\gamma} \\ &\quad \times \left\{ \int_t^u \sqrt{\xi_t(s)} \frac{\xi_t(u)}{(u-s)^\gamma} ds \right. \\ &\quad \left. + \frac{1}{2} \sqrt{\xi_t(u)} \int_u^T \frac{\xi_t(s)}{(s-u)^\gamma} ds \right\} \\ &= \rho^2 \eta^2 2H \int_0^T \sqrt{\xi_0(t)} dt \int_t^T \frac{du}{(u-t)^\gamma} \\ &\quad \times \left\{ \int_t^u \sqrt{\xi_0(s)} \frac{\xi_0(u)}{(u-s)^\gamma} ds \right. \\ &\quad \left. + \frac{1}{2} \sqrt{\xi_0(u)} \int_u^T \frac{\xi_0(s)}{(s-u)^\gamma} ds \right\} + \mathcal{O}(\eta^4). \end{aligned} \quad (\text{B3})$$

For any given initial forward variance curve $\xi_0(u)$, the above expressions for $C^{x\xi}$, $C^{\xi\xi}$ and C^μ may be easily computed numerically.

Appendix 3. Approximate variance of VIX

Let $y_u = \log v_u$ and consider the following approximation of the arithmetic mean by the geometric mean:

$$\begin{aligned}\psi(T) &= \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}[v_u | \mathcal{F}_T] du \\ &\approx \exp \left\{ \frac{1}{\Delta} \int_T^{T+\Delta} \mathbb{E}[y_u | \mathcal{F}_T] du \right\}.\end{aligned}$$

Apart from \mathcal{F}_t measurable terms (abbreviated as ‘drift’), we have

$$\begin{aligned}\int_T^{T+\Delta} E[y_u | \mathcal{F}_T] du &= \eta \sqrt{2H} \int_t^T \frac{dW_s}{(u-s)^\gamma} du + \text{drift} \\ &= \eta \sqrt{2H} \int_t^T \int_T^{T+\Delta} \frac{du}{(u-s)^\gamma} dW_s + \text{drift} \\ &= \eta \frac{\sqrt{2H}}{1-\gamma} \int_t^T \left[(T+\Delta-s)^{1-\gamma} - (T-s)^{1-\gamma} \right] dW_s + \text{drift}.\end{aligned}$$

This gives

$$\begin{aligned}\text{Var}[\log \psi(T) | \mathcal{F}_t] &\approx \frac{\eta^2 D_H^2}{\Delta^2} \int_t^T \left[(T+\Delta-s)^{1/2+H} - (T-s)^{1/2+H} \right]^2 ds \\ &= \eta^2 (T-t)^{2H} f^H \left(\frac{\Delta}{T-t} \right)\end{aligned}$$

where

$$D_H = \frac{\sqrt{2H}}{H+1/2}$$

and

$$f^H(\theta) = \frac{D_H^2}{\theta^2} \int_0^1 \left[(1+\theta-x)^{1/2+H} - (1-x)^{1/2+H} \right]^2 dx.$$

To compute this integral explicitly, we use that, for $\kappa = 1/2 + H$,

$$\begin{aligned}\int_t^T (T+\Delta-s)^\kappa (T-s)^\kappa ds &= \int_0^\tau (s+\Delta)^\kappa s^\kappa ds \\ &= \frac{\tau(\Delta\tau)^\kappa}{1+\kappa} {}_2F_1 \left(-(H+1/2), 3/2+H, 5/2+H, -\tau/\Delta \right).\end{aligned}$$

Thus, we get

$$\begin{aligned}\text{Var}[\log \psi(T) | \mathcal{F}_t] &\approx \frac{2H\eta^2}{\Delta^2} \left[\frac{(\tau+\Delta)^{2(H+1)} - \Delta^{2(H+1)} + \tau^{2(H+1)}}{2(H+1)} \right. \\ &\quad \left. - 2 \frac{\tau(\tau\Delta)^{H+1/2}}{H+3/2} {}_2F_1 \left(-(H+\frac{1}{2}), H+\frac{3}{2}, H+\frac{5}{2}, -\frac{\tau}{\Delta} \right) \right]\end{aligned}$$

with $\tau := T-t$.