



Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets¹

Sigurd Emil Rømer

To cite this article: Sigurd Emil Rømer (2022): Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets¹, Quantitative Finance, DOI: [10.1080/14697688.2022.2081592](https://doi.org/10.1080/14697688.2022.2081592)

To link to this article: <https://doi.org/10.1080/14697688.2022.2081592>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 16 Jun 2022.



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



Article views: 359



View related articles [↗](#)



View Crossmark data [↗](#)

Empirical analysis of rough and classical stochastic volatility models to the SPX and VIX markets¹

SIGURD EMIL RØMER *

Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark

(Received 14 October 2020; accepted 3 January 2022; published online 16 June 2022)

We conduct an empirical analysis of rough and classical stochastic volatility models to the SPX and VIX options markets. Our analysis focusses primarily on calibration quality and is split in two parts. In part one, we perform a historical calibration to SPX options over the years 2004–2019 of a selection of models that include the one-factor rough Bergomi and rough Heston models. In part two, we consider three calibration dates with low, typical, and high volatility, examine a wide selection of models, and calibrate to both SPX options as well as jointly to SPX and VIX options. The key results are as follows: The rough Bergomi and rough Heston models fail to create a term structure of smile effect that is sufficiently pronounced for SPX options. Moreover, we discover that short-expiry SPX smiles generally are more symmetric than long-expiry smiles, a feature we neither find that these models can reproduce. We propose an alternative volatility model driven by two Ornstein-Uhlenbeck processes that uses a non-standard transformation function. Calibrating it to SPX options, we obtain almost perfect fits, and calibrating it jointly to SPX and VIX options, we obtain very decent fits. This suggests, contrary to what one might be led to believe based on much of the existing literature, that the joint SPX-VIX calibration problem is largely solvable with classical two-factor volatility, all without roughness and jumps.

Keywords: Rough volatility; Multifactor volatility; Calibration; SPX options; VIX options

JEL Classifications: C4, C5, C6

1. Introduction

In Gatheral *et al.* (2018) it is argued that volatility on a large number of financial assets is rough. A stochastic process is considered rough if the trajectories are less Hölder continuous than those of Brownian motion. Modelling volatility rough in practice implies explosive volatility paths and an at-the-money implied volatility skew that for many one-factor† models is of the form $\sim cT^{H-\frac{1}{2}}$ where T is the expiry, $H \in (0, \frac{1}{2})$ the so-called Hurst exponent, and c is an anonymous constant. The power-law term structure of skew is remarkably consistent with typical SPX volatility surfaces where often $H \approx 0.1$. Several papers have by now also demonstrated that even simple one-factor rough volatility models can obtain excellent fits to the entire surface, all without jumps; see Bayer *et al.* (2016) for the rough Bergomi model or El Euch

et al. (2019b) for a rough Heston model. Theoretical works on rough volatility are Bayer *et al.* (2019, 2020), El Euch *et al.* (2019a), Forde *et al.* (2021a), Forde *et al.* (2021b), Forde and Zhang (2017), Friz *et al.* (2021), Friz *et al.* (2022), Fukasawa (2017) and Gulisashvili (2021).

The modelling and numerical treatment of volatility derivatives has recently also received attention in the rough volatility literature; see Alòs *et al.* (2022), Gatheral *et al.* (2020), Horvath *et al.* (2020), Jacquier *et al.* (2018), Lacombe *et al.* (2021) and Rømer (2022). An important volatility derivative is the VIX option (a European call or put on the VIX index). As the VIX index is defined in terms of quoted SPX options,‡ the markets for SPX and VIX options are deeply connected. Despite this, the analysis of Chung *et al.* (2011) shows that each market also contains some amount of distinct risk-neutral information and that predictive performance can be improved by taking advantage of both information sets. It is therefore desirable to look for models that can calibrate jointly to SPX and VIX options, something

*Corresponding author. Email: sigurdroemer@gmail.com

¹Some preliminary findings were presented on November 14, 2019, at the *Frontiers in Quantitative Finance* workshop in Copenhagen, Denmark.

† Unless stated otherwise, we by the number of ‘factors’ refer to the number of Brownian motions that drive volatility.

‡ We refer to the VIX whitepaper downloadable via <https://www.cboe.com/education/research> (accessed July 27, 2021).

which is just starting to get explored from the rough volatility perspective.

For the joint problem an often encountered challenge is that of reconciling the typically very steep short-term SPX skews, which in a continuous-path setting imply a very high volatility-of-volatility, with the level of the VIX implied volatility that comparatively is often much lower. In Jacquier *et al.* (2018) where the rough Bergomi model is calibrated, the following is also remarked in relation to the volatility-of-volatility parameter ν : ‘Interestingly, we observe a 20% difference between the parameter ν obtained through VIX calibration and the one obtained through SPX. This suggests that the volatility of volatility in the SPX market is 20% higher when compared to VIX, revealing potential data inconsistencies (arbitrage?)’. As noted in Guyon (2020), the fact that a model does not calibrate well to observed prices need not imply arbitrage, only that it is inconsistent with the market. The joint problem has been described with more rigour in precisely (Guyon 2020) where Julien Guyon shows that an inversion-of-convex-ordering property must be satisfied to fit both markets. As explained in the paper, it for a continuous-path model, in essence, boils down to a volatility process with very fast mean-reversion, and a large, in absolute value, and arguably negative, correlation to the S&P 500 index. The often very steep short term SPX skews suggest that this should be combined with a high volatility-of-volatility.

Previous attempts with continuous-path but classical models have had some success, although, the fits are not always perfect; the double CEV model introduced in Gatheral (2008) is one example. Most (more or less) successful attempts have instead involved jumps; we refer to e.g. the models of Cont and Kokholm (2013), Kokholm and Stisen (2015) and Pacati *et al.* (2018). The inclusion of jumps, however, may be undesirable as it leads to indeterminacy in the construction of hedge portfolios. Continuous-path models are therefore often preferred. The rough volatility finding is promising in this regard since such models naturally produce very fast mean-reversion and high volatility-of-volatility, which suggests that they would do better than classical models in solving the joint problem without jumps. An important related work is Gatheral *et al.* (2020) where the authors calibrate their quadratic rough Heston model and obtain a decent fit to both markets; they though only consider short-term options and use a simplified parameterisation.[†]

Despite the many promising results, we believe the empirical literature is lacking when it comes to the testing of rough volatility pricing models. For example, while several calibrations to SPX options have been published, no systematic validation of the fits exist and we therefore believe it is still an open question if simpler one-factor models such as rough Bergomi and rough Heston truly are consistent with the SPX volatility surface, also across market scenarios. Moreover, we note that only a small selection of rough models have been tested for the joint problem. The present author is in fact only aware of the calibration in Gatheral *et al.* (2020)

of the quadratic rough Heston model (but only to short-term options) and the skewed rough Bergomi model calibrated in Guyon (2018); the latter though fails to reconcile the two markets.

The purpose of our paper is to help fill the gap of empirical work. Our analysis is divided into two parts and focuses primarily on calibration quality. However, a severely limiting factor when calibrating rough models is the slow computational speeds (relative to classical models) that result from the inherent path dependence. Related to this a series of recent papers (Bayer and Stemper 2018, McGhee 2018, Horvath *et al.* 2021) have shown how a pricing model can be significantly sped up with a neural network representation. In part one of our analysis, we therefore employ neural network techniques to calibrate a selection of models to SPX options over the years 2004–2019.[‡] We include the rough Bergomi and rough Heston models, an extended rough Bergomi model, and add the original Heston model of Heston (1993) as a benchmark. The reader will see that these models are incapable of perfect fits across all market scenarios. In search of better results, and to explore the joint problem, we provide a second analysis where we test other mostly more advanced models, including several two-factor volatility models, and a quadratic rough Heston model. We calibrate the models with Monte Carlo to SPX options and jointly to SPX and VIX options, albeit on a smaller set of dates, and analyse the results.

The paper is structured as follows: sections 2–4 cover the first part of our analysis. We define the relevant models in section 2, and, in section 3, outline our neural network methodology and validate the approximations. In section 4, we present and discuss the historical SPX calibrations. The second part of our analysis is provided in section 5 and we conclude in section 6.

2. Setup and models

We consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})$ where \mathbb{Q} a risk-neutral pricing measure and the filtration $(\mathcal{F}_t)_{t \geq 0}$ is generated by two possibly correlated Brownian motions $W_{1,t}$ and $W_{2,t}$, and if not already adapted also by a stochastic process V_t that we call the *instantaneous variance*. Time is measured in years and we always let the filtration be augmented so the usual hypothesis holds. We will write $E_t(\cdot) := E(\cdot | \mathcal{F}_t)$ for conditional expectations. The S&P 500 index is assumed to have the risk-neutral dynamics

$$dS_t = S_t(r(t) - q(t)) dt + S_t \sqrt{V_t} dW_{1,t} \quad (1)$$

where $r(t)$ is the deterministic risk free interest rate, $q(t)$ the continuously compounded proportional deterministic dividend yield. We will write $F_{t,T} = S_t \exp(\int_t^T (r(s) - q(s)) ds)$, $0 \leq t \leq T$, and also $\xi_t(u) = E_t(V_u)$, $0 \leq t \leq u$, for the forward variances. Given a European call or put option with strike K , expiry T , as observed at time t , we define its *log-moneyness* by

[†] The reader should also be aware of the recent developments in using optimal transport theory to solve the joint calibration problem. We refer to the paper (Guo *et al.* 2022) and the references therein but do not discuss the ideas further in this work.

[‡] The trained networks, including interfaces for Matlab, R, and Python, and other codes related to their construction, are made public at https://github.com/sigurdroemer/rough_volatility.

$k = \log(K/F_{t,T})$ and write $\sigma_{BS}(k, T - t)$ for the Black-Scholes implied volatility. The arbitrage free price at time zero is

$$e^{-\int_0^T r(s) ds} E(\max\{\beta \cdot (S_T - K), 0\}) \quad (2)$$

where $\beta = 1$ for a call option, $\beta = -1$ for a put. By an application of Ito's lemma and some simple rewritings, we may also write (2) as

$$F_{0,T} e^{-\int_0^T r(s) ds} E\left(\max\left\{\beta \cdot (\tilde{S}_T - \tilde{K}), 0\right\}\right) \quad (3)$$

where $\tilde{K} = \frac{K}{F_{0,T}}$ and $d\tilde{S}_t = \tilde{S}_t \sqrt{V_t} dW_{1,t}$, $\tilde{S}_0 = 1$. From (3) we see that it suffices to train our neural networks under the assumption of an initial asset price of 1 and zero interest rates and dividends which we will exploit. The general case is recovered by adjusting the strike and rescaling the price.

In what follows, we outline the volatility models that we test in part one of our analysis.

2.1. Heston

In the Heston model of Heston (1993), we have

$$dV_t = \kappa(v_\infty - V_t) dt + \eta \sqrt{V_t} dW_{2,t} \quad (4)$$

where $V_0, \kappa, v_\infty, \eta \geq 0$, $dW_{1,t} dW_{2,t} = \rho dt$, $\rho \in [-1, 1]$. The model has become popular mostly because the characteristic function of $\log(S_t)$ is known analytically which enables fast pricing with Fourier methods. We use a Fourier based pricing formula from Lewis (2001) as it appears in Lord and Kahl (2007), though, rewritten as an integral over \mathbb{R}_+ and with optimal dampening.

2.2. Rough Heston

Different rough Heston models exist in the literature. We focus one that effectively is that of El Euch *et al.* (2019b) which in turn is that from El Euch and Rosenbaum (2018) viewed in the limit where there is zero mean-reversion coming from the drift. More precisely, we consider the model

$$V_t = \xi_0(t) + \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} v \sqrt{V_s} dW_{2,s}, \quad t \geq 0, \quad (5)$$

where $v \geq 0$, $H \in (0, \frac{1}{2})$, $dW_{1,t} dW_{2,t} = \rho dt$, $\rho \in [-1, 1]$. Roughness is generated by the kernel $(t-s)^{H-\frac{1}{2}}$ with lower values of H amplifying the effect. To ensure the existence of a continuous non-negative solution, we assume that

$$\xi_0(t) = V_0 + \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} \psi(s) ds, \quad t \geq 0, \quad (6)$$

where $V_0 \geq 0$ and $\psi \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R})$ are so $\psi(t) dt + V_0 L(dt)$ defines a non-negative measure for $L(dt) = \Gamma(\frac{1}{2} - H)^{-1} t^{-H-\frac{1}{2}} dt$. By example 2.2 and theorem A.2 of Abi Jaber and El Euch (2019a) and table 1 of Abi Jaber *et al.* (2019) it guarantees that (5) has a continuous non-negative solution.

While ξ_0 can in theory be extracted from quoted variance swaps (which in turn can be derived from the volatility surface, see e.g. Gatheral 2006), we shall treat it as a regular parameter to be calibrated along the others, though, as noted, under the form (6). We will likewise treat ξ_0 as an ordinary parameter for the other rough volatility models that we present in this section and where it also appears explicitly in the model definitions.

The characteristic function of $\log(S_t)$ is known up to solving a fractional Riccati equation.[†] While no analytical solution exists there are by now plenty of numerical schemes; see the Adams scheme of Diethelm *et al.* (2004) as used in El Euch *et al.* (2019b) or Abi Jaber (2019), Abi Jaber and El Euch (2019b) and Callegaro *et al.* (2021). Regardless, the reason we focus on the particular rough Heston model formulated in the above is the very fast rational approximation proposed in Gatheral and Radoičić (2019) for solving the fractional Riccati equation and which is valid for that model. As the authors show, the rational approximation is very accurate for negative ρ and low H , which is also the region relevant to SPX options. To price options, we combine the rational approximation with the same Fourier pricing formula that we use for the classical Heston model.

2.3. Rough Bergomi

The rough Bergomi model of Bayer *et al.* (2016) is defined by

$$V_t = \xi_0(t) \mathcal{E}\left(\eta \sqrt{2H} \int_0^t (t-s)^{H-\frac{1}{2}} dW_{2,s}\right), \quad t \geq 0,$$

where $\eta \geq 0$, $H \in (0, \frac{1}{2})$, $dW_{1,t} dW_{2,t} = \rho dt$, $\rho \in [-1, 1]$, and we have defined $\mathcal{E}(X) := e^{X - \frac{1}{2}E(X^2)}$ for a general mean-zero Gaussian variable X . As V_t is given explicitly, only mild assumptions, including non-negativity, apply to ξ_0 to ensure that also (1) has a valid solution. In practice, we will assume ξ_0 piecewise constant and non-negative which suffices. We price options with Monte Carlo and to that end simulate V_t with the hybrid scheme of Bennedsen *et al.* (2017b) and use the conditional estimator of McCrickerd and Pakkanen (2018) as a base estimate. We use a log-Euler scheme to simulate the part of S_t that is not integrated out by the conditioning and add the expectation of that as a control variate to our price estimator. We use 50 000 paths, half of which are antithetic.

2.4. Extended rough Bergomi

The rough Bergomi model can be extended so the volatility process is driven by two factors with different roughness levels: We define the extended rough, or fractional, Bergomi model by[‡]

$$V_t = \xi_0(t) V_{1,t} V_{2,t}, \quad t \geq 0, \quad (7)$$

[†] See equations (3) and (4) of El Euch *et al.* (2019b) which hold also under our assumptions on ξ_0 as follows by a special case of theorem 2.3 of Abi Jaber and El Euch (2019a).

[‡] The model is not our own invention as we were inspired by a no longer existing GitHub page by Ryan McCrickerd.

where

$$\begin{aligned} V_{1,t} &= \mathcal{E} \left(\zeta \sqrt{2\alpha + 1} \int_0^t (t-s)^\alpha dW_{1,s} \right), \\ V_{2,t} &= \mathcal{E} \left(\lambda \sqrt{2\beta + 1} \int_0^t (t-s)^\beta dW_{2,s} \right), \quad t \geq 0, \end{aligned} \quad (8)$$

and $\alpha, \beta \in (-\frac{1}{2}, \frac{1}{2})$, $\zeta, \lambda \in \mathbb{R}$. For this model, it is assumed that $W_{1,t}$ and $W_{2,t}$ are *independent*. As for the non-extended version, only mild assumptions apply to ξ_0 and having it be piecewise constant and non-negative likewise suffices as we will also assume in practice. When $\alpha = \beta$, the model reduces to the ordinary rough, or fractional, Bergomi model with $H = \alpha + \frac{1}{2}$ and

$$\rho = \frac{\zeta}{\sqrt{\zeta^2 + \lambda^2}}, \quad \eta = \sqrt{\zeta^2 + \lambda^2}. \quad (9)$$

We will use the reparameterisation (9) also when $\alpha \neq \beta$. We may then translate back to the original formulation as $\zeta = \eta\rho$ and $\lambda = \eta\sqrt{1 - \rho^2}$. We also use Monte Carlo to price options under (7)–(8). To this end, we simulate $V_{1,t}$ and $V_{2,t}$ as for rough Bergomi and use similar variance reduction techniques. That is, except for the conditional Monte Carlo method which we could not carry over. We use anywhere between 25 000 and 100 000 paths depending on the shape of ξ_0 ; more paths when volatility is low as we found estimation difficult in those scenarios.

The extended model differs from the original in that the separation of the Hurst exponents $\alpha + \frac{1}{2}$ and $\beta + \frac{1}{2}$ between the correlated and uncorrelated factor, allows for a corresponding separation of the explosion rates for the skew and curvature of the option smiles. To see why, note that

$$\text{Var}(\log V_{1,t}) = \zeta^2 t^{2\alpha+1}, \quad \text{Var}(\log V_{2,t}) = \lambda^2 t^{2\beta+1}, \quad t \geq 0,$$

so for $\alpha > \beta$, the correlated factor $V_{1,t}$ dominates at long time horizons, the uncorrelated factor $V_{2,t}$ at short horizons (vice-versa if $\alpha < \beta$). As we illustrate in figure 1 for $\alpha > \beta$, the effect on option prices is as claimed. Compared to what

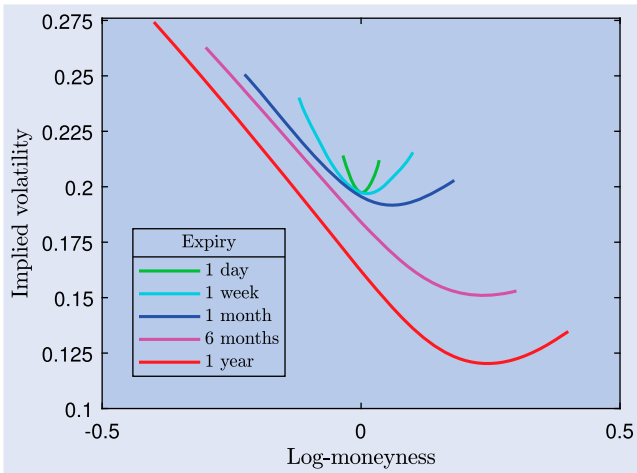


Figure 1. Volatility smiles under extended rough Bergomi with $\alpha = 0.4$, $\beta = -0.4$, $\rho = -0.9$, $\eta = 2.1$, $\xi_0(t) = 0.20^2$.

the original model would produce, the option smiles are now more symmetric at short expiries and get increasingly skewed at longer horizons. The opposite would be observed if $\alpha < \beta$. The reader will see that the SPX volatility surface often have decoupled term structures for the skew and curvature which is why we include the model.

3. Neural network approximations

In what follows, we explain how we build neural network representations of the pricing models. We start by introducing neural networks, then review some existing approaches on applying them to option pricing, and explain our own adaptation. Lastly, we validate the trained networks.

3.1. Neural networks

While neural networks are helpful for many purposes, we will introduce and motivate them in the context of approximating some multivariate function $F : \mathbb{R}^N \rightarrow \mathbb{R}^M$. We imagine that F is slow to evaluate and that we seek to replace it with a neural network that ideally will be faster but still sufficiently accurate. To this end, we consider a fully connected feed-forward neural network (henceforth just a 'neural network'). This is a mapping $F_{\text{net}} : \mathbb{R}^N \rightarrow \mathbb{R}^M$ defined by its general architecture and a set of weights denoted w ; a network architecture is shown in figure 2. To evaluate F_{net} , one traverses from left to right in the graph, and at each node, the input layer excluded, performs a computation of the form $x \mapsto \sigma(a^\top x + b)$ where x is the output from the last layer (a column vector), a a coefficient column vector, b a scalar, and σ an (activation) function. The weights w are the collection of all coefficients a and b . The problem of finding a neural network approximation to F consists of finding an architecture and weights w so the error between F_{net} and F is small.

The weights are typically chosen as follows for a fixed architecture: First we generate a synthetic 'training' dataset of input-output pairs by evaluating F at various inputs x and storing the results $y = F(x)$. The inputs are often sampled from some distribution covering the relevant domain of F . For a random input sample X , we define the *generalisation error* by the number

$$E[\mathcal{L}(F_{\text{net}}(X; w), F(X))] \quad (10)$$

where $\mathcal{L} : \mathbb{R}^M \times \mathbb{R}^M \rightarrow \mathbb{R}$ is a loss function and $F_{\text{net}}(X; w)$ denotes the neural network evaluated in X with weights w . An empirical version of (10) can be constructed on the training data by averaging errors across the samples. To train the network, one can minimise the empirical version with respect to w . This is usually performed with stochastic gradient descent which is a gradient-based optimisation method. For the method one iterates in epochs, i.e. cycles, across the training data, first shuffling all samples, then looping over smaller batches, covering all of them, each time updating the weights with a gradient estimate. The final error is typically evaluated on a 'test' dataset that is generated independently

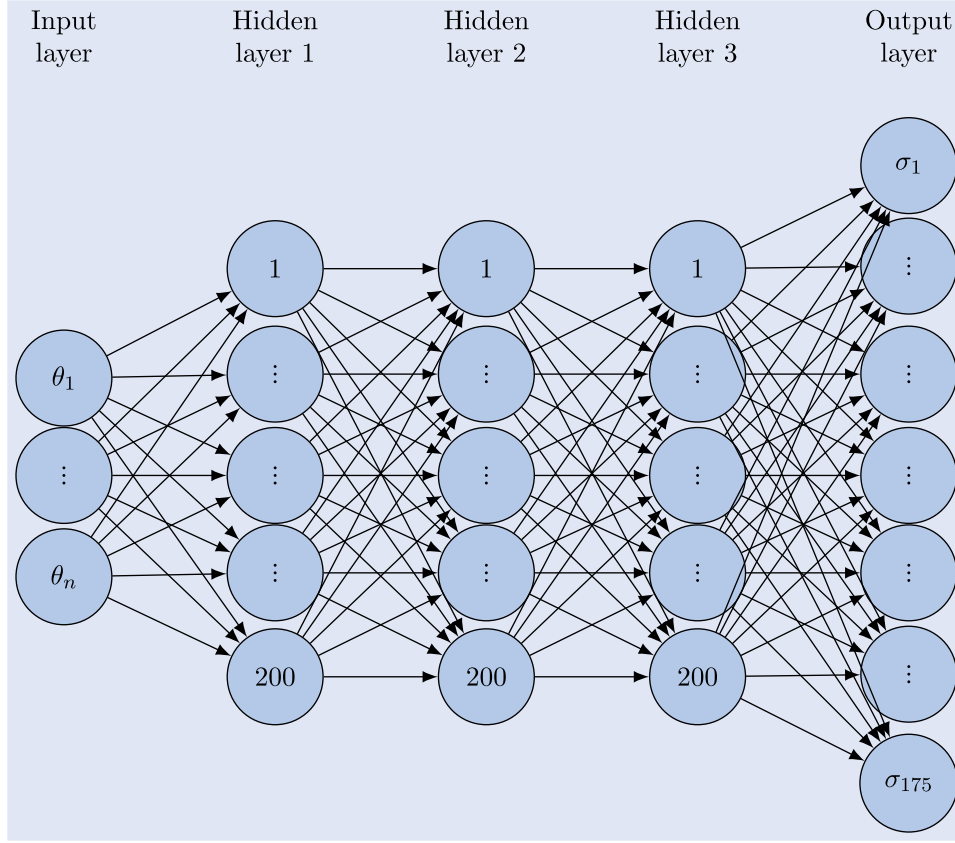


Figure 2. Our neural network design for implied volatilities on expiry slice in the interval $[0, 0.008]$ for a model with n inputs parameters. Parameters are denoted θ_i for $i = 1, \dots, n$ and implied volatilities σ_i for $i = 1, \dots, 175$.

of the training data. We refer to Bishop (2006) and Goodfellow *et al.* (2016) for the finer details in the training of neural networks.

The literature contains many results that justify neural networks for function approximation. The Universal Approximation Theorem of Hornik (1991) for example essentially states that a continuous function can be approximated to any desired precision on a compact domain using networks of a bounded number of layers but arbitrarily many neurons (nodes) per layer. A dual result can be found in Kidger and Lyons (2020) for networks with a bounded number of neurons but arbitrary many layers. A common observation in an option pricing context is nevertheless that only a few hidden layers are worthwhile. The authors of Bayer and Stemper (2018) e.g. report that their approximations did not consistently improve when going beyond four hidden layers. This is likely because the mapping from parameters to prices tends to be smooth so there is little benefit in introducing complex non-linearities beyond that achievable with a few layers.

3.2. Neural networks for options pricing

In the following, we formulate the approximation problem in relation to options pricing, calibration in particular, and review existing approaches. We will frame the situation very abstractly and consider therefore a general pricing model with some parameter space $\Theta \subset \mathbb{R}^{d_1}$, $d_1 \in \mathbb{N}$, and let $\Lambda \subset \mathbb{R}^{d_2}$, $d_2 \in \mathbb{N}$, denote a space of contract parameters such as strike-expiry pairs for a number $N_c \in \mathbb{N}$ of financial contracts. We

consider a pricing function $P : \Theta \times \Lambda \rightarrow \mathbb{R}^{N_c}$ that maps from model and contract parameters to a vector of prices which could be expressed in currency units or implied volatility. It is such a function P that we wish to approximate with a neural network.

Example: Under Heston we may set $\Theta = \mathbb{R}_+^4 \times [-1, 1]$ and for $\theta \in \Theta$ write $\theta = (V_0, \kappa, v_\infty, \eta, \rho)$ in our previous notation. If we consider European calls and puts, we could set $\Lambda = \mathbb{R}_+^2$ and let $\lambda = (K, T) \in \Lambda$ denote the strike and expiry of a given option. Then $P(\theta, \lambda) = P(V_0, \kappa, v_\infty, \eta, \rho, K, T)$ would be its price. Alternatively, we could consider a fixed set of, say, $N_c = 10$ options specified as strike-expiry pairs $(K_i, T_i) \in \mathbb{R}_+^2$, $i = 1, \dots, 10$, set $\Lambda = \emptyset$, and let $P : \Theta \rightarrow \mathbb{R}^{10}$ return a vector with their prices.

The general setup and how we choose to formulate P can have important consequences for the approximations and whether or not interpolation or extrapolation is needed in addition. In the following, we therefore discuss advantages and disadvantages of different neural network setups and definitions of P . All cases reviewed concern the pricing of European call and put options, and P always return prices in implied volatility; the same will be true in our own setup.

We comment first on the work (Bayer and Stemper 2018) where a *pointwise* learning approach is used. In the paper, the authors set $\Lambda = \mathbb{R}_+^2$ and let an element of this space refer to the strike and expiry of a single European call or put option. Their pricing function $P : \Theta \times \Lambda \rightarrow \mathbb{R}$ then returns the implied volatility of such a single option. They use a network with 4 hidden layers, each with 4096 neurons, which

gives roughly 67 million network weights to train. To focus the accuracy of the network on the relevant parts of the contract space, they sample strikes and expiries for the training data from a joint distribution constructed from observed SPX contracts.

In Horvath *et al.* (2021) an *image-based implicit* learning approach is proposed. For their numerical examples, the authors set $\Lambda = \emptyset$ and let $P : \Theta \rightarrow \mathbb{R}^{88}$ be the pricing function that returns implied volatilities for options on the 11 by 8 strike-expiry grid defined by the Cartesian product below (it is here assumed that $S_0 = 1$):

$$\begin{aligned} &\{0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5\} \\ &\times \{0.1, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8, 2.0\}. \end{aligned} \quad (11)$$

They use a network with 4 hidden layers, each containing 30 neurons, which results in about 7000 network weights. To price contracts that are not in the grid, appropriate interpolation (and extrapolation) must be used. The method of course generalises to other contract grids. It is called *image-based* as the network learns the mapping from model parameters to several points on the volatility surface ($N_c > 1$). The word *implicit* refers to the fact that the pricing function depends implicitly on the chosen contracts ($\Lambda = \emptyset$).

The authors of Horvath *et al.* (2021) argue that their method has a number of advantages. We point out a few as we see it: (1) The complexity is reduced as we only need to learn the volatility surface on a fixed set of points. This should facilitate the use of smaller networks. (2) While training, information on nearby contracts are considered jointly which allows the weights to adjust to the entire surface at once. This could speed up the training process. (3) The reduction in the input dimension and the increased information contained in an image likely implies fewer training samples are needed. This is advantageous when combined with Monte Carlo as one can reuse simulated paths for multiple contracts in the grid.

A downside compared to the pointwise method is the need for at least interpolation which adds more errors and computational costs. However, in a calibration setting, the contracts to evaluate remain fixed as points in \mathbb{R}_+^2 . Consequently, we have in this context found it possible to perform pre-computations to speed up repeated evaluation of the interpolation scheme.[†] Since we neither had problems controlling the interpolation error, we decided to use the image-based method ourselves, although with a few adjustments due to a problem that we shall now explain.

More precisely, we find that it can be problematic to use a Cartesian product of contracts in the strike-expiry space such as (11). To explain the issue, say e.g. that we estimate P with Monte Carlo. We then argue that we should only include contracts in the grid that, with an acceptable amount of paths, can be accurately estimated. Indeed, if samples with significant errors enter into the training dataset this will likely have a negative influence on the accuracy and reliability of the trained network. We should therefore avoid options that are too far out-of-the-money. What is ‘far out-of-the-money’ is however highly dependent on the strike, expiry, and model parameters.

[†] We were at least able to do so with our choice of interpolation method which we explain in a moment.

Consequently, with a grid in the form of a Cartesian product, one is forced to use relatively narrow strike bounds to control the Monte Carlo error across expiries and parameters.

One solution is to use a dynamic strike range. That is, we could look for functions $\tilde{K}_{\min}, \tilde{K}_{\max} : \Theta \times \mathbb{R}_+ \rightarrow \mathbb{R}$ that satisfy $\tilde{K}_{\min}(\theta, T) < \tilde{K}_{\max}(\theta, T)$ for all $(\theta, T) \in \Theta \times \mathbb{R}_+$ and which should specify strike bounds covering a realistic range of values where θ is a parameter vector, T an expiry. Such a technique is used in McGhee (2018) where neural networks are trained for the SABR model of Hagan *et al.* (2002). There are nevertheless a few challenges: (1) The strike bounds of McGhee (2018) use intricate properties of the SABR model and it is not obvious how to generalise them to other models. (2) The calibration task requires us to price all observed contracts, possibly after filtering, for any model parameter in the domain of optimisation. By construction this is not in general possible with a dynamic moneyness range. That is, unless a robust extrapolation technique can be used, but we do not believe this is a trivial matter.

To balance the discussed advantages and disadvantages, we have for our own project chosen a middle-of-the-road solution in that we use the image-based method, and that with dynamic moneyness bounds, but only allowing the bounds to depend on the expiry. In this way we can pre-filter the market data to the support of our networks, and yet, to some degree, accommodate how the probability distribution of the underlying asset changes across the different time horizons.

We show our choice of bounds in figure 3. To specify the contracts for the image-based method, we consider the below 64 expiries

$$\begin{aligned} \mathcal{T} := \{ &0.002, 0.003, \dots, 0.01, 0.015, \dots, 0.05, 0.06, \dots, \\ &0.20, 0.225, 0.25, 0.30, \dots, 0.5, 0.6, \dots, 3\}, \end{aligned}$$

and use 25 uniformly spaced points in the log-moneyness dimension between the bounds. This results in 1600 contracts.

In McGhee (2018) it was suggested to partition the volatility surface into different expiry groups and to train separate networks for each. We ourselves find that this can help improve the approximations and keep the networks of a manageable size. A likely reason is that short- and long-term

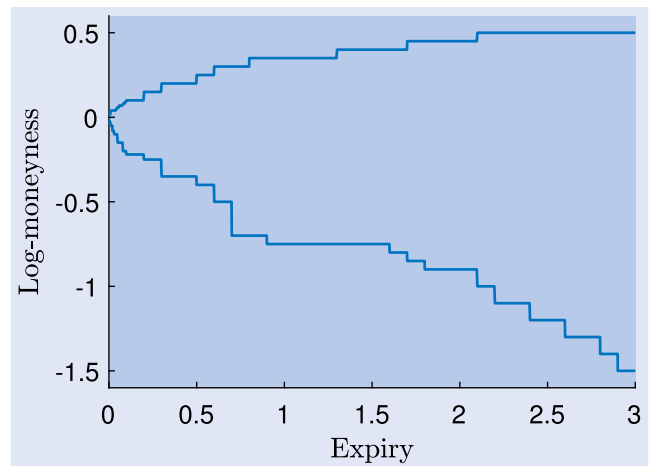


Figure 3. Moneyness region supported by our neural networks.

volatility smiles behave rather differently, especially under rough volatility, which suggest that they should be handled separately. We train six networks per model, each of which covers expiries in the intervals $[0, 0.008]$, $(0.008, 0.03]$, $(0.03, 0.12]$, $(0.12, 0.4]$, $(0.4, 1]$, and $(1, 3]$.

Lastly, let us explain how we interpolate the volatility surface: Say that we wish to obtain the implied volatility of a contract with log-moneyness k and expiry T assumed within the domain of figure 3. We then interpolate as follows:

- Find the closest expiries $T_1, T_2 \in \mathcal{T}$ so $T_1 \leq T \leq T_2$
- Interpolate with a natural cubic spline the volatility smiles for the expiries T_1 and T_2 to obtain $\sigma_{BS}(k, T_1)$ and $\sigma_{BS}(k, T_2)$
- Apply linear interpolation between $\sigma_{BS}^2(k, T_1)T_1$ and $\sigma_{BS}^2(k, T_2)T_2$ in the time dimension to obtain $\sigma_{BS}^2(k, T)T$ and thereby $\sigma_{BS}(k, T)$

Our natural cubic spline interpolation does not rule out static arbitrage. However, as follows by the work of Roper (2010), a necessary condition for no static arbitrage is that $T \mapsto \sigma_{BS}^2(k, T)T$ is non-decreasing. Thus if implied volatilities are arbitrage free at the points (k, T_1) and (k, T_2) , our choice of time-interpolation ensures that this is not violated.

An unreported numerical test showed that our interpolation method mostly ensures an absolute relative error in implied volatility below 10^{-3} .

3.3. The forward variance curve

We now explain how we handle ξ_0 numerically and sample it for the datasets. Due to the special assumptions that apply under rough Heston, we handle it somewhat different between the models.

For the rough Bergomi models, we parameterise ξ_0 piecewise constant as

$$\xi_0(t) = \sum_{i=1}^n 1_{(t_{i-1} \leq t < t_i)} \bar{\xi}_i, \quad 0 \leq t < t_n, \quad (12)$$

where $n \in \mathbb{N}$, $\{t_i\}_{i=0}^n \subset \mathbb{R}_+$, $0 = t_0 < t_1 < \dots < t_{n-1} < t_n$, $\bar{\xi} := \{\bar{\xi}_1, \dots, \bar{\xi}_n\} \subset \mathbb{R}_+$. We extend $\xi_0(t) = \bar{\xi}_n$ for $t \geq t_n$. In practice, we set $n = 27$ and let $\{t_i\}_{i=0}^n$ be the 28 time points below:

$$\{0, 0.0025, 0.005, \dots, 0.02, 0.04, \dots, 0.16, 0.28, \dots, 1, 1.25, 1.5, 1.75, 2, 3\}. \quad (13)$$

The grid corresponds roughly to daily sections for the first week, then weekly sections till the first two months, monthly sections for the remainder of the first year, quarterly sections for the second year, and one section for the third year.

Let us then discuss how to sample realistic forward variance curves for the datasets under (12). In Horvath *et al.* (2021) ξ_0 was likewise assumed piecewise constant but with 8 flat sections each of which was sampled i.i.d. uniformly distributed. While it is tempting to use i.i.d. sampling in our context also, we note that it is unlikely to adequately cover the space of possible curves due to the fine grid that we use.

To explain why, consider for simplicity the equidistant case where $t_i = T \frac{i}{n}$, $i = 0, 1, \dots, n$, for some $T > 0$. Then if we sample $\bar{\xi}_1, \dots, \bar{\xi}_n$, i.i.d., the variance swap quote with maturity T , which we denote by $VS(T)$, by the law of large numbers, converges to a fixed value as we subdivide the curve into more independent sections (i.e. as $n \rightarrow \infty$). More precisely:

$$VS(T) := \frac{1}{T} \int_0^T \xi_0(s) ds = \frac{1}{n} \sum_{i=1}^n \bar{\xi}_i \rightarrow E(\bar{\xi}_1), \quad n \rightarrow \infty.$$

Indeed, the variability in the sampled variance swap quote has disappeared. We can generally expect similar problems with the sampling variability for a finite but large n (relative to T), not-necessarily equidistant discretisations, and when looking at variance swaps at other maturities. The key to improving the sampling variability is to break the independence assumption when we sample adjacent sections of the forward variance curve. In what follows, we outline one idea to this end.

Consider the Heston model where

$$VS(T) = v_\infty + (V_0 - v_\infty) \frac{1 - e^{-\kappa T}}{\kappa T}, \quad T > 0. \quad (14)$$

One solution is then to sample the Heston parameters κ, v_∞, V_0 , and for the given sample to compute $\bar{\xi}$ as finite-difference derivatives of $T \mapsto VS(T)T$ under (14). To help the networks generalise to curves that are not based on the parametric form (14) one can add some smaller i.i.d. noise on top. We use this idea to sample most of our curves, although the technical details of our implementation are somewhat more involved. To add more variety we also sample other types of curves including i.i.d. ones as in Horvath *et al.* (2021). We refer to the publicly available code for the details. We shall denote the distribution from which we sample $\bar{\xi}$ by D_ξ^{rb} .

For the rough Heston model, we, as noted, first and foremost assume that ξ_0 is parameterised of the form (6) in terms of (V_0, ψ) assuming also that $\psi(t) dt + V_0 L(dt)$ defines a non-negative measure.[†] For our implementation, we further assume ψ piecewise constant as

$$\psi(t) = \sum_{i=1}^n 1_{(t_{i-1} \leq t < t_i)} \bar{\psi}_i, \quad 0 \leq t < t_n, \quad (15)$$

where $n \in \mathbb{N}$, $\{t_i\}_{i=0}^n \subset \mathbb{R}_+$, $0 = t_0 < t_1 < \dots < t_{n-1} < t_n$, $\bar{\psi} := \{\bar{\psi}_1, \dots, \bar{\psi}_n\} \subset \mathbb{R}$. We extend the curve flat as $\psi(t) = \bar{\psi}_n$ for $t \geq t_n$. The measure $\psi(t) dt + V_0 L(dt)$ is non-negative under our setup if

$$\bar{\psi}_i \geq -\frac{V_0}{\Gamma(\frac{1}{2} - H)} t_i^{-H-\frac{1}{2}}, \quad i = 1, \dots, n-1, \quad (16)$$

and $\bar{\psi}_n \geq 0$ since the density of $L(dt) = \Gamma(\frac{1}{2} - H)^{-1} t^{-H-\frac{1}{2}} dt$ is decreasing in t . Now ξ_0 is parameterised in terms of $(V_0, \bar{\psi})$ more specifically. As it will be more natural, we will reparameterise in terms of a set of variables $\bar{\xi} := \{\bar{\xi}_0, \bar{\xi}_1, \dots, \bar{\xi}_n\} \subset \mathbb{R}_+$

[†] Technically ξ_0 also depends on H , though, we shall consider the Hurst parameter given in this context.

that will be the forward variances with maturities $\{t_i\}_{i=0}^n$. To this end, note that if we insert (15) into (6), we obtain, at the grid points,

$$\xi_0(t_i) = V_0 + \sum_{k=1}^i w_{i,k} \bar{\psi}_k, \quad i = 0, 1, \dots, n,$$

where we interpret the sum as zero when $i = 0$ and have defined

$$w_{i,k} := \frac{1}{\Gamma(H + \frac{3}{2})} \left((t_i - t_{k-1})^{H+\frac{1}{2}} - (t_i - t_k)^{H+\frac{1}{2}} \right),$$

$$1 \leq k \leq i \leq n.$$

Define then

$$\bar{\xi}_i := V_0 + \sum_{k=1}^i w_{i,k} \bar{\psi}_k, \quad i = 0, 1, \dots, n,$$

so $\bar{\xi}$ precisely are the forward variances with maturities $\{t_i\}_{i=0}^n$. The variables $\bar{\xi}$ constitute a valid reparameterisation as we may uniquely recover $(V_0, \bar{\psi})$ iteratively from the equations

$$\bar{\psi}_i = \frac{\bar{\xi}_i - V_0 - \sum_{k=1}^{i-1} w_{i,k} \bar{\psi}_k}{w_{i,i}}, \quad i = 1, \dots, n,$$

and $V_0 = \bar{\xi}_0$. For our rough Heston implementation, we also set $n = 27$ and let $\{t_i\}_{i=0}^n$ be given by (13). To sample $\bar{\xi}$ for the datasets, we take first as given a sample of H . We then keep resampling $\bar{\xi}$ similarly to how we do it for the rough Bergomi models until we get a sample where both (16) and $\bar{\psi}_n \geq 0$ holds. The reader is again referred to the publicly available code for the details. We denote the distribution from which we, conditionally on H , sample $\bar{\xi}$ for rough Heston by $D_{\bar{\xi}|H}^{\text{rh}}$.

3.4. Sampling the test and training data

In what follows, we explain how we sample and compute the datasets: Prices are computed under the assumption of $S_0 = 1$ and $r(t) = q(t) = 0$ which as discussed suffices. For each model, we generated 136 000 samples for the training dataset and an additional 24 000 samples for a test dataset which we use to validate the final approximation error. This is double the amounts used in Horvath *et al.* (2021) for the rough Bergomi model. In table 1, we show our sampling distributions (distr.) for all models and parameters (par.). By \mathcal{U} we denote the uniform distribution. Columns a and b signify the sampling bounds. For the rough Bergomi models, we found that we could sample H , respectively β , from an only approximately uniform distribution and thereby roughly halve the total computation time. We use the symbol \mathcal{U}^* to refer to this distribution, though, the reader must consult the code for the details. To sample a parameter vector for a given model, we sample each individual parameter independently from the distributions shown in the table, though, with the following nuances and exceptions: (1) As noted, we sample $\bar{\xi}$ under rough Heston *conditionally* on H . (2) We sample all parameters except $\bar{\xi}$ with the Sobol sequence. Since

Table 1. Parameter sampling distributions.

Heston				Rough Heston			
Par.	Distr.	a	b	Par.	Distr.	a	b
κ	\mathcal{U}	0	25	H	\mathcal{U}	0	0.5
η	\mathcal{U}	0	10	ν	\mathcal{U}	0.10	1.25
ρ	\mathcal{U}	-1	0	$\bar{\rho}$	\mathcal{U}	-1	0
$\sqrt{V_0}$	\mathcal{U}	0.05	1	$\bar{\xi}$	$D_{\bar{\xi} H}^{\text{rh}}$	0.05 ²	1
$\sqrt{V_\infty}$	\mathcal{U}	0.05	1				
Rough Bergomi				Extended rough Bergomi			
Par.	Distr.	a	b	Par.	Distr.	a	b
H	\mathcal{U}^*	0	0.5	α	\mathcal{U}	-0.5	0.5
η	\mathcal{U}	0.75	3.5	β	\mathcal{U}^*	-0.5	0.5
ρ	\mathcal{U}	-1	0	η	\mathcal{U}	0.75	3.5
$\bar{\xi}$	$D_{\bar{\xi}}^{\text{rb}}$	0.05 ²	1	ρ	\mathcal{U}	-1	0
				$\bar{\xi}$	$D_{\bar{\xi}}^{\text{rb}}$	0.075 ²	1

quasi-random numbers are often more regularly spaced than pseudo-random numbers, the generalisation error (10) should then be better estimated by its empirical version. We expect this to lower the tendency of overfitting to the training data. (3) To ensure high-quality datasets, we discarded samples where option prices were not convex across strikes, where Merton's tunnel was violated, or where numerical integration for the Heston models failed to converge. For each discarded sample we generated a replacement.[†]

3.5. Hardware and software

The training and test datasets were computed on a server running Intel Xeon Platinum 8175 3.1 GHz CPUs with 384 GB of RAM and 48 physical cores (96 logical). It took around a week to finish them all. The neural networks were thereafter trained in Python. For rough Heston we used Python 3.6.1 with the Keras 2.4.3 library and Tensorflow 2.3.4 as the backend. For the other models we used Python 3.7.1 with the Keras 2.2.4 library and TensorFlow 1.13.1 as the backend. Computational times reported in the remainder of the paper are recorded in Matlab 2019a on a laptop that runs a 1.6–3.4 GHz Intel Core i5 8250U CPU with 4 cores (8 logical processors) and 8 GB of RAM.

3.6. Hyperparameters

By hyperparameter optimisation we refer to the act of minimising the generalisation error by tuning all other aspects of a neural network and its training than the weights themselves. Examples of such aspects are the number of layers, the number of neurons per layer, the choice of and settings

[†] The lower volatility bound is set to 7.5% for the extended rough Bergomi model and to 5% for the others. The reason is that despite using 100 000 paths in the case of low volatility, we for the extended model found it particularly difficult to estimate prices with sufficient accuracy for volatilities below 7.5%. If we did not adjust the sampling bound we would have discarded too many samples to finish the datasets in a reasonable amount of time.

Table 2. Distribution of absolute relative errors between the trained networks and the test datasets measured in implied volatility.

Model	50th	75th	95th	99th	99.9th	99.99th	Max.
Heston	0.09%	0.16%	0.51%	1.40%	4.97%	15.63%	219.82%
Rough Heston	0.08%	0.19%	0.63%	1.54%	4.60%	11.39%	130.05%
Rough Bergomi	0.18%	0.37%	1.03%	2.44%	7.52%	18.19%	165.17%
Extended rough Bergomi	0.40%	0.70%	1.38%	2.47%	5.69%	11.17%	51.18%

Note: Numbers in the header represent percentiles.

Table 3. Distribution of relative standard errors on the training datasets measured in implied volatility.

Model	50th	75th	95th	99th	99.9th	99.99th	Max.
Rough Bergomi	0.13%	0.23%	0.44%	0.85%	2.06%	3.98%	278.51%
Extended rough Bergomi	0.51%	0.60%	0.72%	1.14%	2.47%	4.63%	256.49%

Note: Numbers in the header represent percentiles.

for the optimiser, the batch size, and the choice of activation and loss function. Hyperparameter optimisation is very costly as it requires one to retrain the network across many different configurations. We therefore fixed most hyperparameters based on common choices found in the literature that generally appear to work well: We choose 3 hidden layers as used in a preprint version of Horvath *et al.* (2021), the Adam algorithm of Kingma and Ba (2015) as our optimiser, Elu as our activation function except for a linear output layer, and root-mean-squared-error as the loss function. We scale the inputs and outputs as in Horvath *et al.* (2021) and use 200 neurons per layer which we found to be sufficient. For the training, we first let the optimiser run for 500 epochs with a batch size of 32 at which point both the training and test losses stabilised. However, by increasing the batch size to 5000 and continuing for another 200 epochs, we found that we could reduce the error by an additional moderate amount.[†] The final errors on the test datasets turned out similar to those on the training datasets which indicates that overfitting was not a problem.

3.7. Accuracy and speed

We examine the accuracy and speed of our networks. Throughout we measure the accuracy in terms of implied volatility. In table 2 we show percentiles of the absolute relative errors between the trained networks and the test datasets. We note that the errors generally are below a few per cent and mostly within a single per cent. In table 3 we for the rough Bergomi models show the relative standard errors; by this we refer to the standard errors of implied volatility divided by the estimated implied volatility. We see that also these errors are comfortably low. The errors versus the test datasets are generally larger than the standard errors which is unsurprising. On our filtered SPX dataset, the 25th and 75th percentiles of the relative bid-ask spreads are 1.36% and 4.25%; we here refer to the bid-ask spread measured in implied volatility divided by

[†] We hypothesise that a batch size of 32 may allow for fast convergence initially, but as one nears a minimum, the gradient estimates may be too noisy for full convergence.

the implied volatility of the mid quote. The approximations are mostly within the bid-ask spreads and we are therefore confident in using our networks for the calibrations.

In terms of speed, we find that we can evaluate all networks for a given model once in about $7.1 \cdot 10^{-4}$ seconds (average of 10 000 runs under rough Bergomi). When we calibrate, we however need to compute prices many times on a scattered set of points anywhere in the contract domain. This will cost more time as we will need to interpolate. As noted, the points are though fixed and we therefore find it possible to perform pre-computations to speed up repeated evaluation of the interpolation scheme. If we fix the 4166 contracts we have available on May 15, 2019, perform the pre-computations, and evaluate both the neural networks and the interpolation scheme 10 000 times, we obtain for rough Bergomi an average running time of $1.8 \cdot 10^{-3}$ seconds per evaluation. The cost of interpolation is therefore small when spread across many evaluations.

4. Historical calibration to SPX options

In this section, we present part one of our empirical work. We use a dataset obtained from <https://datashop.cboe.com> which consists of bid and ask quotes on SPX European call and put options on trading days between May 3, 2004, and May 15, 2019. The quotes are recorded at 15:45 Eastern time, and we apply a number of filters to remove unreasonable and low liquidity quotes and to ensure a minimum number of strikes and expiries on each date. Interest rates and dividends are implied from the available put-call parities. In-the-money contracts and those outside the domain of figure 3 are removed. After filtering, the dataset consists of 4 643 880 bid-ask pairs spread over 3775 trading days which corresponds to an average of 1230 observations per day.

To state the calibration problem, consider a given trading day and let σ_i^{bid} , σ_i^{ask} , σ_i^{mid} , denote the implied volatility of the bid, ask, and mid quote for the i 'th observed option. Similarly, let $\sigma_i^{model}(\theta)$ denote the implied volatility of the option under the given model with parameter vector θ . We then define the

weighted root-mean-square-error (wRMSE) by

$$\text{wRMSE}(\theta) := \sqrt{\sum_i w_i (\sigma_i^{\text{mid}} - \sigma_i^{\text{model}}(\theta))^2} \quad (17)$$

where the sum is over all contracts from the given date and w_i are weights that are normalised so $\sum_i w_i = 1$. We use the weights to make the error measure more robust. There is e.g. a considerable amount of time variation in the dataset in that the amount and location of the quoted expiries changes a lot from 2004 to 2019. We therefore allocate 15 % of the weights to expiries within 1 month, 35 % to expiries between 1 and 6 months, and the rest to longer expiries. Additionally, we allocate the same total weight to each distinct expiry within each expiry group. To account for differences in liquidity, we further, up to normalisation, weigh each contract by the factor

$$\frac{1}{0.01 + \sigma_i^{\text{ask}} - \sigma_i^{\text{bid}}}.$$

To calibrate Heston, we minimise the wRMSE. For the rough models, we often find that $\bar{\xi}$ is overparameterised which can result in calibrated curves that look unrealistic. The main culprit is that there may be multiple grid points for the curve between each observed expiry. For the rough Bergomi models, we therefore use the below algorithm to merge $\bar{\xi}$ -parameters before calibrating:

Merging $\bar{\xi}$ -parameters: Loop $i = 1, 2, \dots, n - 1$: If there are no expiries observed in $(t_{i-1}, t_i]$ or (t_i, ∞) , we merge $\bar{\xi}_i = \bar{\xi}_{i+1}$.

We use the same algorithm before we calibrate rough Heston, except we keep $\bar{\xi}$ -values that correspond to forward variance maturities at or below 0.28 unmerged. The reason for the latter is that we find merging too many of the parameters in the short end of the curve can significantly limit the flexibility in fitting the volatility term structure as we for this model also have to satisfy the non-negativity requirements (16) and $\bar{\psi}_n \geq 0$ which we also impose when we calibrate.

Unfortunately, even merging $\bar{\xi}$ -parameters as explained, we yet find example dates where the calibrated curves look unrealistic. Besides not merging parameters for the short end of the curve under rough Heston, another possible explanation is noise coming from the bid-ask spread. We therefore additionally use penalisation when we calibrate and minimise instead $\text{wRMSE}(\theta)^2 + \lambda C(\bar{\xi})$ with respect to θ where $\lambda \in \mathbb{R}_+$ and $C(\bar{\xi}) := \sum_{i=1}^n (\sqrt{\bar{\xi}_i} - \sqrt{\bar{\xi}_{i-1}})^2$ measures the non-smoothness of $\bar{\xi}$; we here set $\bar{\xi}_0 = \bar{\xi}_1$ for the rough Bergomi models by convention. We find that $\lambda = 0.002$ works well without worsening the fits too much. Calibrating the models across the entire dataset results in average differences of only 4–5 bps (basis points) in wRMSE between using $\lambda = 0.002$ and $\lambda = 0$. We are therefore comfortable using $\lambda = 0.002$ for our analysis.[†]

[†] All optimisations are performed in Matlab with either the *lsqnonlin* function with the *trust-region-reflective* algorithm or the *fmincon* function with the *sqp* algorithm. For rough Heston, we always run first an optimisation without the non-negativity constraints (16) and $\bar{\psi}_n \geq 0$ which generally is faster and often returns a solution that satisfy them regardless. If the solution does not satisfy the constraints, we run a second optimisation with them imposed.

4.1. Calibration results

We present the calibration results in figures 4–6, though, we have left out most of the Heston parameters for brevity. Daily closing values of the VIX index are provided for reference in figure 4 (bottom).[‡] The reader should note that for a model of the form (1), the VIX index squared is, under mild conditions, essentially defined by

$$\text{VIX}_t^2 = 100^2 \left[\frac{1}{\Delta} \int_t^{t+\Delta} \xi_t(u) du \right], \quad t \geq 0, \Delta = \frac{1}{12}, \quad (18)$$

and thus, up to scale, represents the risk-neutral expected average spot variance over the next month; see e.g. Gatheral (2006) for details.

We start with a discussion of the calibration errors which we show in figure 4 as centralised 20-day moving averages. The first thing we note is that the rough models outperform classical Heston substantially overall. While the average error is around 80 bps for Heston, it is close to 50 bps for the rough models. The error for Heston also deteriorates rapidly around the financial crisis of 2008–2009, remains elevated for several years thereafter, and at one point even reaches values as high as 160 bps. In contrast, the errors for the rough volatility models mostly stay below 80 bps and also appear less sensitive to the volatility level. In terms of the latter, note that we for these models observe average errors in the ranges 52–60, 51–55, and 46–50 bps, when the VIX index respectively is in the ranges 0–15, 15–25, and above 25. Under Heston the average errors are 69, 80 and 100 bps for the same VIX levels. It follows also that the relative errors for, at minimum, the rough models generally are larger when volatility is low. Evidence from sections 4.3 and 5 suggest that the volatility surface displays an increasingly complex structure at lower volatility levels. We believe this, at least partly, explains it. Another possible explanation is that we due to our fixed mon-eyness bounds can expect fewer contracts to be filtered when volatility is low.

While the classical Heston model clearly performs worst, it is less obvious which of the rough models is better. If we compare rough Bergomi and rough Heston, we obtain an average difference of 6 bps in favour of rough Heston, though, this is dwarfed by the standard deviation of the differences which is 11 bps. Likewise, pitting the extended and ordinary rough Bergomi models against each other, we compute an average difference of only 1 bps and a standard deviation of 4 bps. It therefore appears that there are no noteworthy improvements to be made in choosing between the tested one-factor rough models and that there neither are any notable gains to be made in separating the Hurst exponents between the two independent Brownian motions that drive volatility. This, however, does not mean that the fits are perfect or that one cannot construct better models. Additional results presented later will make this clear.

In the following, we comment on figures 5–6. Because ρ and η are effectively equal between the rough Bergomi

[‡] A single observation is missing from our VIX index dataset and is therefore excluded from any figures and calculations; the dataset is downloaded from https://www.cboe.com/tradable_products/vix/vix_historical_data on May 30, 2021.

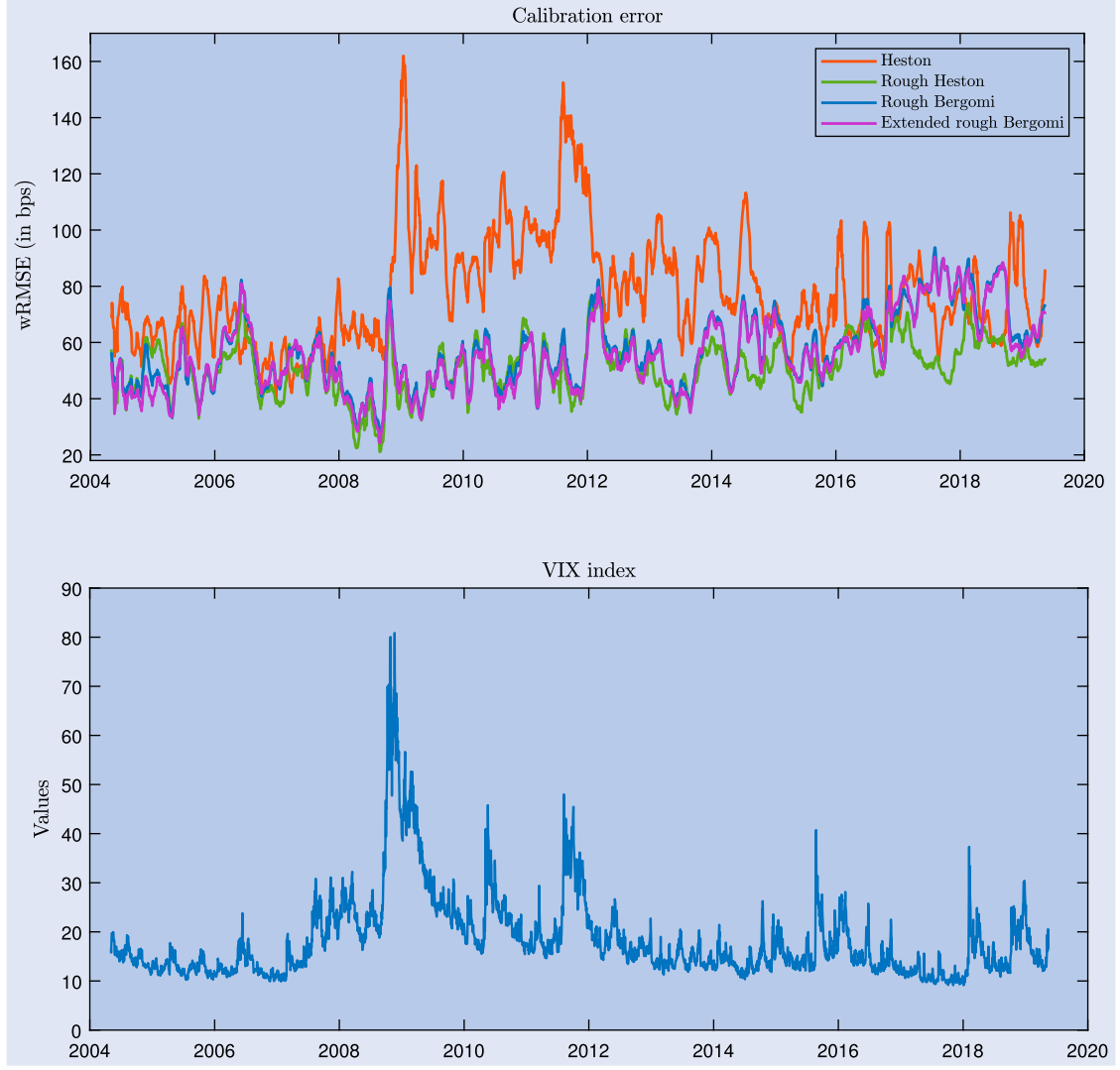


Figure 4. Calibration errors shown as 20-day centralised moving averages and VIX closing values.

models as calibrated, we leave out the extended model when we comment on these parameters, though, not for the Hurst exponents where a non-trivial observation can be made.

We start with the correlation parameters. Note first how ρ for both Heston models mostly fluctuate in the rather narrow range from -0.8 to -0.6 and stay a good distance away from the lower bound of -1 . This is in contrast to ρ under rough Bergomi which sometimes calibrates very close to -1 and whose time evolution is more unstable; the standard deviation of daily changes in ρ is 0.027 for rough Heston while it is 0.042 for rough Bergomi. In terms of the level, we believe extreme values such as $\rho \approx -1$ suggest unrealistic dynamics for the index price as it implies that the volatility surface is driven by essentially a single underlying factor. If we compare with figure 4, we also see that the error often deteriorates relative to rough Heston when ρ is near -1 . This is especially true around the years 2017–2018 where the gap widens considerably to 20–30 bps. All of this hints at some structural problems with the rough Bergomi model.

With regards to the Hurst exponents, we observe a strong relation between the values calibrated for rough Bergomi and rough Heston. This is hardly surprising given that they both

control the explosion rate of at-the-money skew.[†] In addition, with values mostly below 0.25 across the entire time series, and averages at respectively 0.09 and 0.10 , it is clear that option prices, as viewed through these models, consistently suggest that volatility is very rough. Interestingly enough, the Hurst exponents correlate positively with volatility; the correlation between H and VIX is 0.51 under rough Bergomi and 0.10 under rough Heston. That volatility is less rough in periods of high volatility is a finding that also appears in Bennedsen *et al.* (2021) under the realised measure. Because H controls the speed at which volatility's dependence on the history of $W_{2,t}$ dissipates, with lower H implying a faster dissipation, it suggests that the SPX market implies weaker mean reversion when volatility is high. We believe this is meaningful from an economic point of view: typical high volatility events (think e.g. of the financial crisis) arguably tend to influence markets for longer than events that correspond to smaller volatility increases.

[†] We refer to Bayer *et al.* (2016) for a result saying that the at-the-money implied volatility skew is of the form $\sim cT^{H-\frac{1}{2}}$ under rough Bergomi and to remark 1.8 of Bayer *et al.* (2020) for the same under rough Heston.

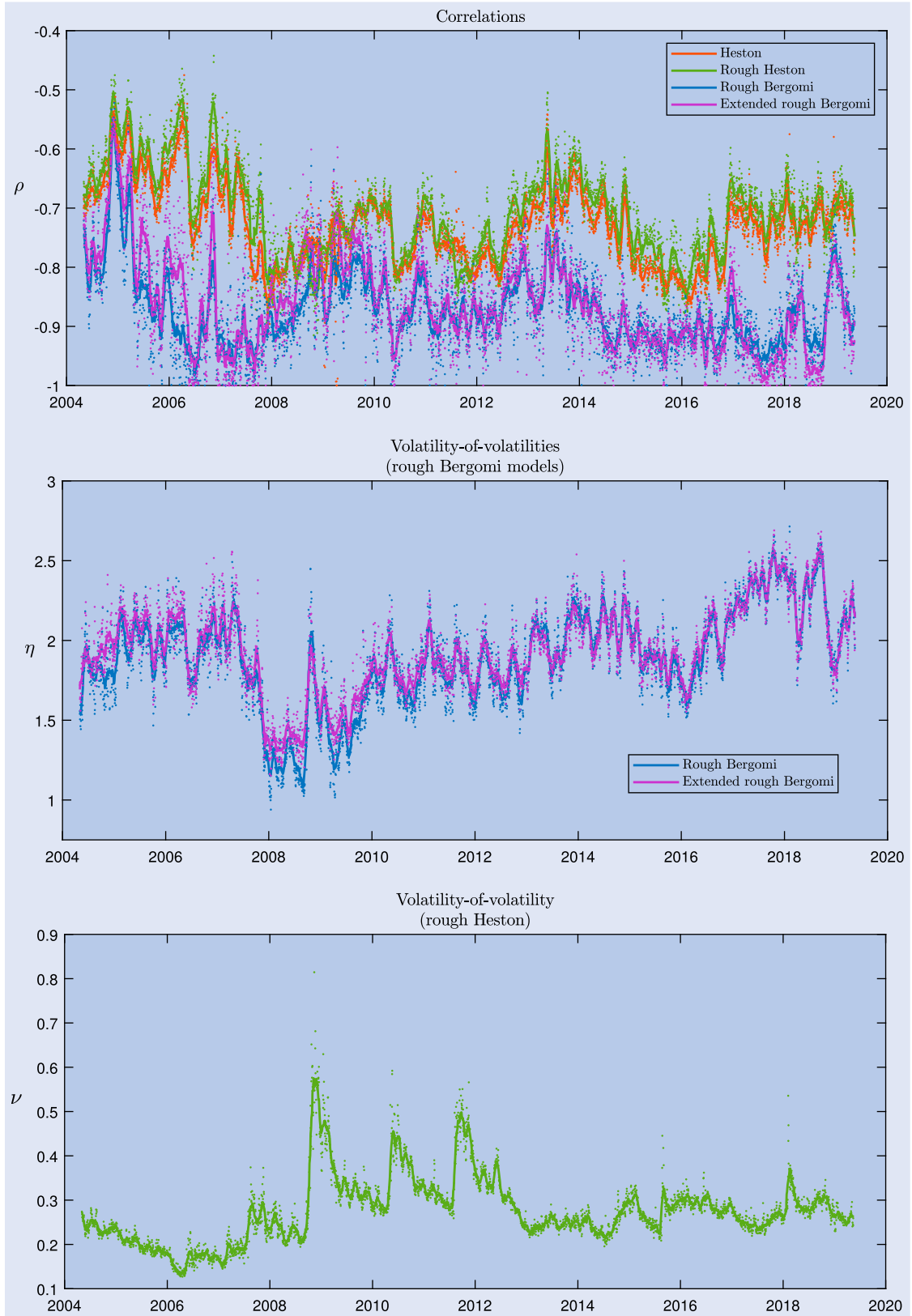


Figure 5. Calibrated correlation and volatility-of-volatility parameters. Solid lines show centralised 20-day moving averages. Dots show individual observations.

We could also interpret the calibrated H -values as reflecting the volatility smoothness implied by the market: By theorem A.1 of Abi Jaber and El Euch (2019a) and example 2.3 of Abi Jaber *et al.* (2019) volatility is locally Hölder continuous of all

orders less than H for the rough Heston and rough Bergomi models, that is, up to discontinuities in ξ_0 for the latter. This is a consequence of the rate of singular decay of the kernel function, or in other words the speed of ‘mean-reversion’

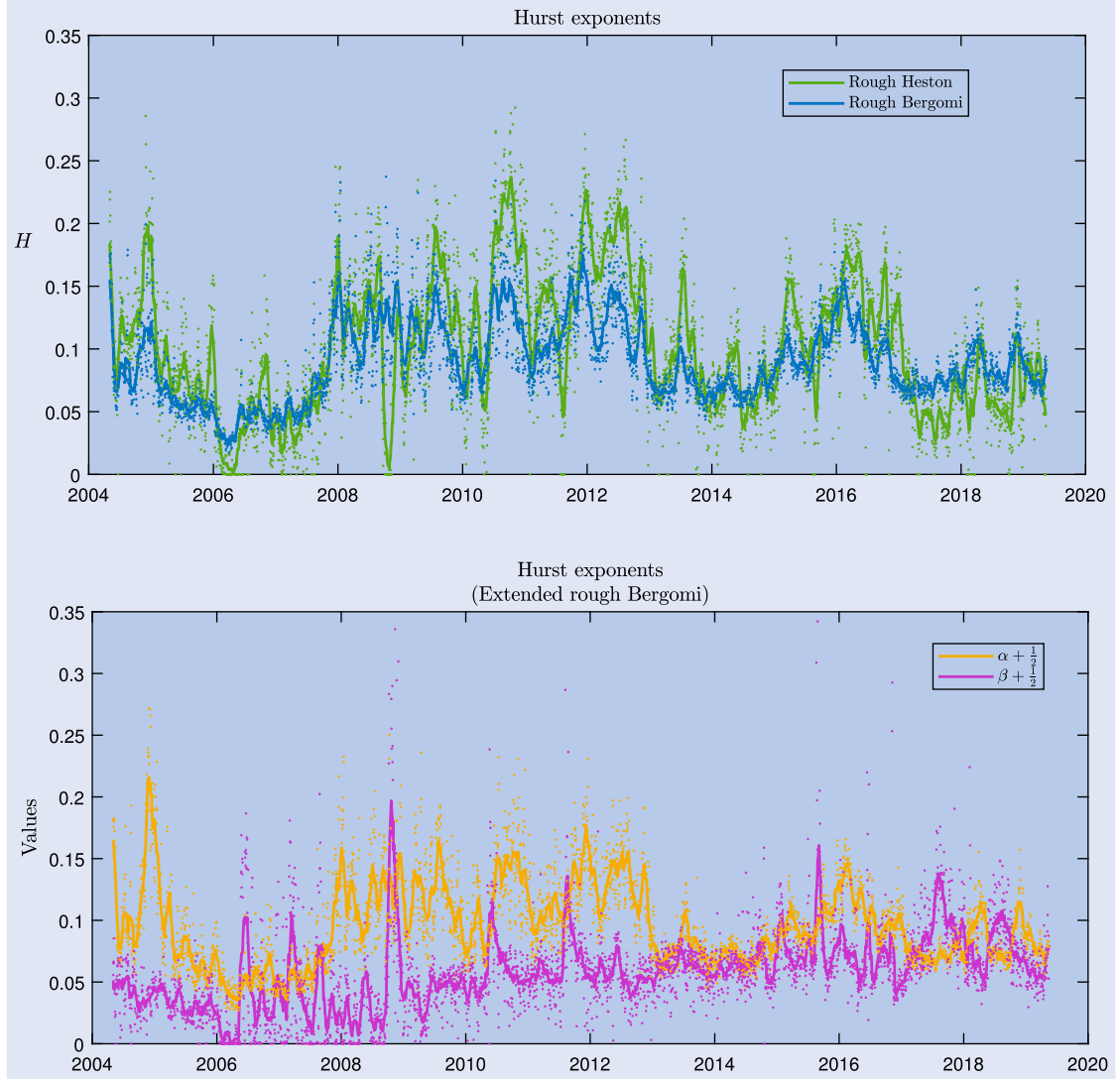


Figure 6. Calibrated Hurst exponents. Solid lines show centralised 20-day moving averages. Dots show individual observations; a few are left out to improve visibility.

across very small time-scales.[†] It is then possible that a deeper explanation for the level dependence could be found by examining market microstructure models. We e.g. note that El Euch *et al.* (2018) shows a connection between certain features of high frequency trading and roughness. Nevertheless, a critique of interpreting the calibrated H -values as representing the implied volatility smoothness is that H for the fractional kernel is responsible for ‘mean-reversion’ at both short and long time scales. Thus, unless prices truly are generated by a model with exactly the fractional kernel, the calibrated values are bound to reflect an imperfect trade-off between the mean-reversion implied at short and long time horizons. In fact, in Bennedsen *et al.* (2021) it is suggested to decouple the short and long time behaviour of the volatility autocorrelations.

It should be pointed out that while rough Bergomi sometimes calibrates extreme ρ ’s, rough Heston at times calibrates extreme values of H near zero. The time series of H for rough

Heston also appears less stable than for rough Bergomi; this could also explain why the computed correlation between H and VIX is lower for rough Heston. The extreme values of ρ (rough Bergomi) and H (rough Heston) tend to occur in the same time periods which indicates that, in fact, both models may lack some flexibility and that the problems likely are related.

Consider now α and β under the extended rough Bergomi model: Note that the time series of $\alpha + 1/2$ largely resembles that of H for rough Bergomi while that of $\beta + 1/2$ is somewhat different. The first observation is consistent with the fact that our calibrated rough Bergomi models are very similar. More precisely, for values of ρ near -1 , as we do observe, the parameter $\lambda = \eta\sqrt{1 - \rho^2}$ is small in absolute value compared to $\zeta = \eta\rho$ and therefore, as calibrated, the extended rough Bergomi model is very similar to the ordinary one but with Hurst exponent $\alpha + 1/2$. It follows also that we should be careful in drawing conclusions based on the time series of $\beta + 1/2$ as the effect of β on the extended model is small for such ρ ’s. However, that we with relative consistency observe $\alpha > \beta$ does at least *indicate* that the short-term smiles are more symmetric than the long-term ones. We back this observation up

[†] We generally use the expression ‘mean-reversion’, later also ‘auto-correlation’, liberally as e.g. rough Bergomi is not stationary.

Table 4. Statistics on per day calibration times shown in seconds.

Model	Min.	5th	50th	95th	Max.
Heston	0.04	0.06	0.09	0.15	0.60
Rough Heston	0.26	0.42	1.17	6.52	41.12
Rough Bergomi	0.10	0.17	0.52	1.41	4.30
Extended rough Bergomi	0.08	0.15	0.40	1.35	3.16

Note: Numbers in header represent percentiles.

with more robust evidence in section 4.3. Also, as the reader will see in section 5, there are better ways to decouple the term structures of skew and curvature.

For the volatility-of-volatilities, note that ν for rough Heston correlates highly and positively with volatility; the correlation between ν and VIX is 0.76. This is evidence that the standard deviation of volatility does not scale enough with volatility. Stated differently, it appears that the term $\sqrt{V_s}$ of (5) is misspecified and should be replaced by V_s^γ for some $\gamma > \frac{1}{2}$. Although the evolution of η for rough Bergomi is more stable, we again observe a state-dependence, except in the other direction; the correlation between η and VIX is -0.47 . As volatility is log-normal under rough Bergomi, we believe this is evidence that γ should be somewhere between $\frac{1}{2}$ and 1, though, probably closer to 1 as the state-dependence is less strong under rough Bergomi.[†] Our observations are qualitatively in line with those of Rømer and Poulsen (2020) where estimations on realised volatility under a one-factor non-rough diffusion model suggest γ is in the range 0.91–0.97.

Lastly, we consider the calibration speed which we show statistics on in table 4. We note that we mostly calibrate in less than a second on what can be considered a standard laptop as of the publication date. This shows that we with neural networks can implement rough volatility in a realistic setting and obtain speeds that are feasible also in a production environment. That rough Heston calibrates somewhat slower overall is likely a consequence of the fact that we impose also the non-negativity constraints (16) and $\tilde{\psi}_n \geq 0$.

4.2. Prediction quality

We examine the predictive quality of the models. To this end, we perform the following experiment: On each trading day, and for each model, we fix the calibrated parameters that are not state variables. We then recalibrate up to 20 trading days into the future, allowing only the state variables to change. For Heston, this means that we only recalibrate the instantaneous variance, and for the rough volatility models it means that we only recalibrate the forward variances. Market variables such as the index price and yields are updated as observed on the given future day.

In figure 7 (left), we show the mean prediction errors at different horizons; standard errors are at or below 1 bps. As could be expected, the rough models beat classical Heston at all horizons and the errors also grow more slowly. Furthermore,

[†] Note that $d\xi_t(u)/\xi_t(u) = \eta\sqrt{2H}(u-t)^{H-\frac{1}{2}}dW_{2,t}$ under rough Bergomi which corresponds to $\gamma = 1$.

while rough Heston on average calibrates better in-sample by 6 bps compared to rough Bergomi, its predictions are almost 10 bps worse at the 20-day horizon. Like with the in-sample calibrations, there neither in this case are any meaningful differences between the extended and ordinary rough Bergomi model.

In figure 7 (right), we show the time evolution of the prediction errors at the 5-day horizon (shown as 20-day centralised moving averages). The plot largely resembles the evolution of the in-sample errors, though, with the key difference that the performance of rough Heston often deteriorates rapidly in periods of high volatility with errors reaching as high as 150 bps. In contrast, the errors under the rough Bergomi models are more stable and rarely go above even 100 bps. This also explains why the predictive quality deteriorates faster in figure 7 for rough Heston. We suspect that the culprit is the elasticity parameter γ which likely is far too low in that model.

4.3. Decoupled term structures of skew and curvature

In what follows, we analyse the in-sample calibrations in more detail. In figure 8, we show the fits for a subset of the expiries on September 4, 2012, and January 13, 2017. The calibration errors on the first date are 64, 44, 51, and 49 bps for, respectively, Heston, rough Heston, rough Bergomi and the extended rough Bergomi model, and, listed in the same order, are 70, 70, 81, and 89 bps on the second date.[‡] If we compare with the top of figure 4, we see that in terms of the errors, and for the rough models, September 4, 2012, represents a typical example, whereas January 13, 2017, represents one of the more problematic dates. In figure 9, we show at-the-money skews computed from the market data. Power-law fits are added for comparison and we have used the definition:

$$\text{Skew}(T) := \left. \frac{\partial \sigma_{BS}(k, T)}{\partial k} \right|_{k=0}, \quad T > 0.$$

We consider first the fits on September 4, 2012, where the term structure of at-the-money skew is reasonably well represented by a power-law consistent with what typical one-factor rough volatility models display. The volatility smiles shown in figure 8 are also fitted relatively well by the rough models on this date, though, improvements can still be made. The Heston model, however, cannot reproduce the power-law term structure of skew—see e.g. the functional form derived in Gatheral (2006, p. 35)—and we believe this is why it results in a worse fit, especially to the very short expiry.

In contrast, on January 13, 2017, even the rough models calibrate quite poorly. The main reason is that the volatility surface, on this date, displays a pronounced term structure of (a)symmetry as the short-term smiles are significantly more symmetric than the long-term ones. Unsurprisingly this feature is not reproduced by the one-factor models rough

[‡] For the latter date, the ordinary rough Bergomi model performs slightly better than the extended one. Section 5 contains similar examples where nested models perform marginally better. We believe it can be explained by smaller differences in the neural network approximations, Monte Carlo error, and/or different initial guesses.

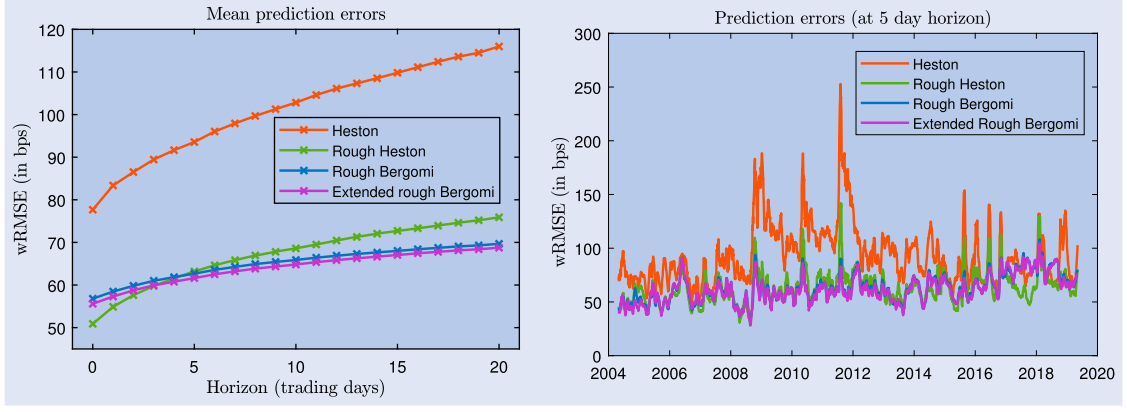


Figure 7. Left: Mean prediction errors at different horizons. Right: Prediction errors at the 5-day horizon (shown as 20-day centralised moving averages). Note that the blue line is mostly hidden behind the magenta coloured one.

Bergomi and rough Heston, though less obviously is neither captured by the extended rough Bergomi model. Figure 9 shows that the term structure of at-the-money skew also far from resembles a power-law. The very short-term skews even flatten as the smiles become increasingly symmetric.[†] On a side-note, we remark that the very short-term smile on September 4, 2012, also is just slightly more symmetric than what our rough volatility models could produce. This hints at the possibility that the term structure of smile (a)symmetry is a structural feature of the volatility surface.

The fits on January 13, 2017, especially, and the associated non-power-law term structure of observed skews, demonstrate that typical one-factor rough volatility models are not always consistent with the SPX volatility surface. A likely explanation is that two volatility factors with different temporal properties and correlations to the index price are required. This view is precisely the reason why we included the extended rough Bergomi model in our analysis, though, as can be observed, it is apparently insufficient to only separate the Hurst exponents for each of the two underlying independent Brownian motions. The results of section 5 will show that better fits can be achieved with proper two-factor volatility, i.e. using three Brownian motions to model (S_t, V_t) .

We provide now a systematic investigation of the rough volatility fits across the entire historical dataset. We exclude the extended rough Bergomi model from our analysis as it calibrates effectively the same as the ordinary one. We then do the following for each trading day and model: First, we fix the Hurst exponent H as already calibrated. Separately for each expiry we then recalibrate the correlation parameter ρ , the volatility-of-volatility parameter either η or ν , and the forward variance curve ξ_0 which here will be assumed flat. In this way, we achieve almost perfect fits to each expiry slice. We can then use the recalibrated correlation and volatility-of-volatility parameters to study in closer detail how the models should be adjusted for a better fit.

We present the results in figure 10, where we for various quantities show percentiles, overall means, and means split by the level of the VIX index. The expiries are placed into groups for the computations and the first axes are log-scaled. The

plots in the top row show the wRMSE before recalibration minus that after.[‡] We note that the recalibrations have generally lowered the errors by a good amount. The improvement is most notable for short and long expiries which is unsurprising as a full calibration will seek to balance both ends. In the second row, we show values for the relative wRMSE before recalibration minus that after. The relative wRMSE is defined by:

$$\text{Relative wRMSE}(\theta) = \sqrt{\sum_i w_i \left(\frac{\sigma_i^{\text{mid}} - \sigma_i^{\text{model}}(\theta)}{\sigma_i^{\text{mid}}} \right)^2}.$$

We note that the fits are improved by a good amount in relative terms also, and we therefore believe that the recalibrations can meaningfully be used for our analysis.

Let ρ be the correlation parameter obtained from the full calibration for a given model on a given day and let $\tilde{\rho}(T)$ denote the recalibrated version for an expiry T . Likewise, let η and ν denote the calibrated volatility-of-volatility parameters and let $\tilde{\eta}(T)$ and $\tilde{\nu}(T)$ denote their recalibrated versions. To structure our analysis, we then look at the ratios $\tilde{\rho}(T)/\rho$, $\tilde{\eta}(T)/\eta$, and $\tilde{\nu}(T)/\nu$, to see in which direction we should shift the distribution of S_T for different time horizons T . The reader can find the results for the ratios in the bottom two rows of figure 10. As can be observed, we for both models, on average, need less correlation (more symmetry) and more volatility-of-volatility (more smile effect) at short expiries, vice versa at long expiries, and all relative to what the models could produce in the full calibration experiment. The observations are statistically significant as the standard errors of the means are all at or below 0.01.

The first observation shows that it is a structural feature of the SPX volatility surface that short expiry smiles are more symmetric. On the second observation, note that a faster decaying volatility autocorrelation function tends to increase the smile effect at short expiries relative to long expiries. See e.g. the expansion formula of Bergomi and Guyon (2012). There it is also noted that ‘...a slow decay of the ATM skew (respectively curvature) can be traced, in the context of a stochastic volatility model, to the existence of

[†] The non-exploding skew on January 13, 2017, though, need not imply that volatility is not rough (recall figure 1).

[‡] For a proper comparison we always rescale the weights w_i so they sum to one for a given expiry.

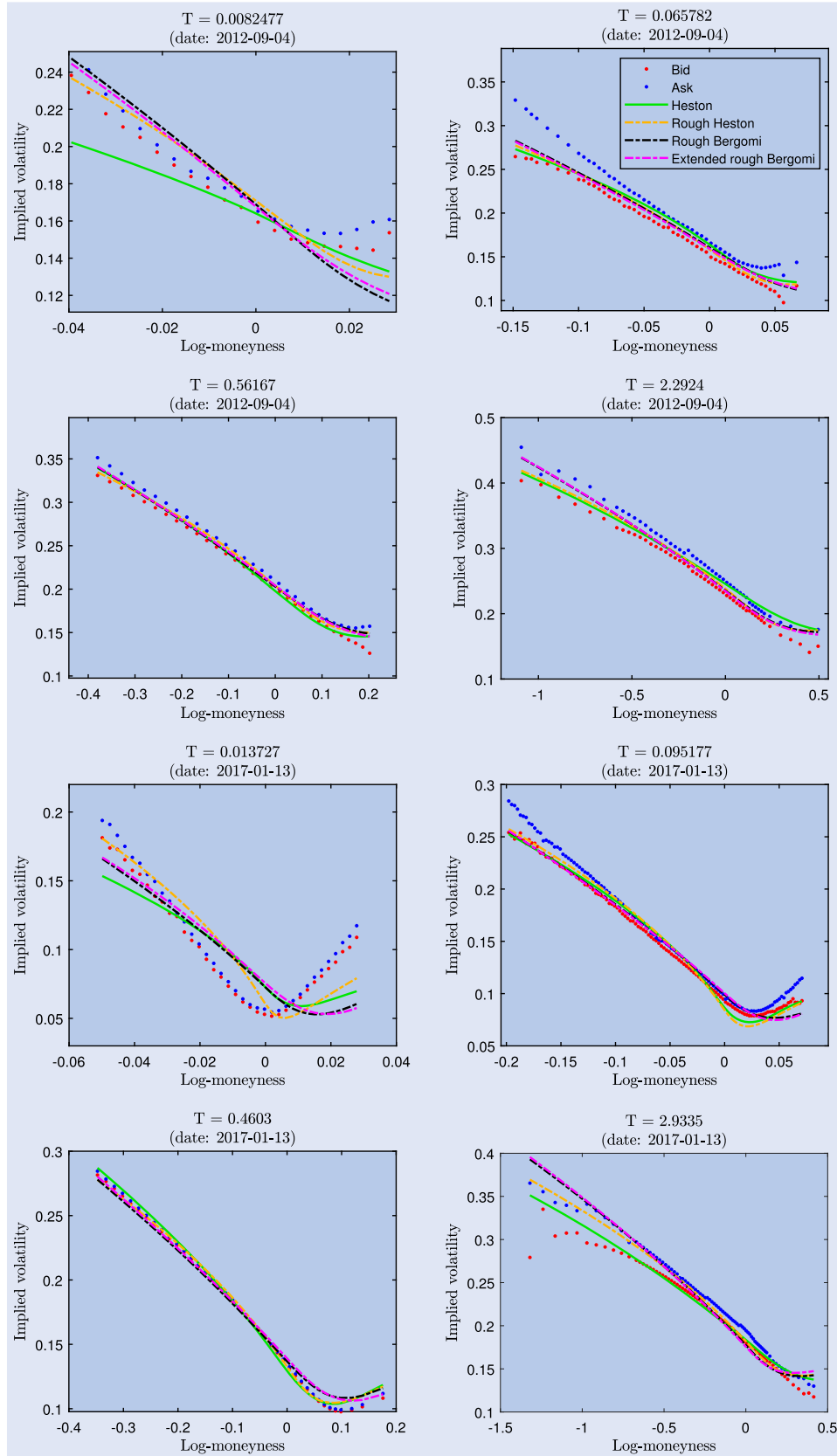


Figure 8. Example calibrations from the historical SPX analysis. Only a subset of the expiries are shown.

long-range spot/variance (respectively variance/variance) correlations...[†] A natural conclusion is then that the volatility

[†] That $\text{Skew}(T) \approx cT^{H-\frac{1}{2}}$ for many one-factor rough models is just one example of how ‘mean-reversion’ and thus presumably a

autocorrelations implied by the calibrated rough Bergomi and

stronger separation of short and long lag autocorrelations, here controlled by H , creates a more pronounced term structure of smile effect. As the January 13, 2017, smiles exemplify, we should though be careful to interpret the term structure of skew directly as a measure

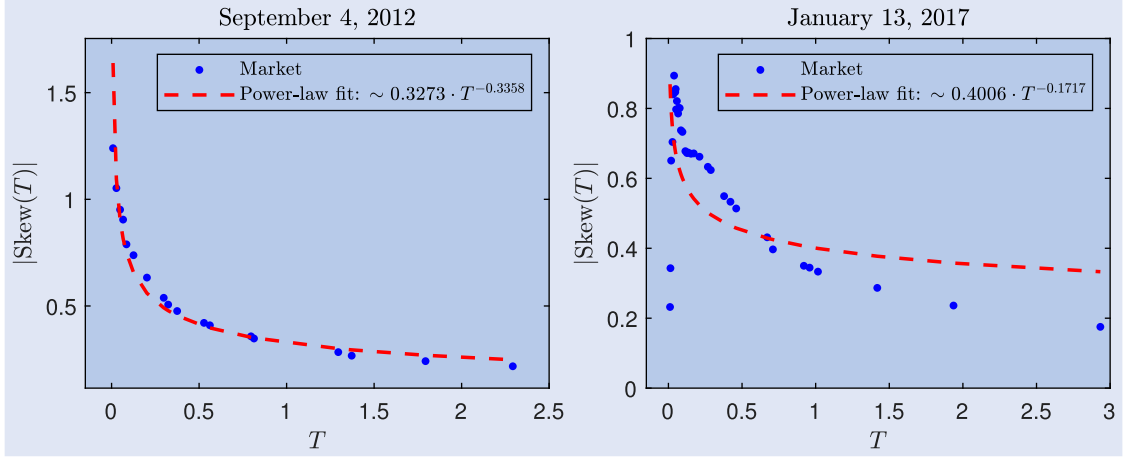


Figure 9. At-the-money skews. Market skews are extracted by fitting the SVI parameterisation of Gatheral and Jacquier (2014) to each volatility smile. Power-laws are fitted by log-linear regressions on the market skews.

rough Heston models fail to sufficiently separate the short and long time-scale properties of volatility. We can think of two immediate possible reasons for this: Either the fractional kernel does not have enough flexibility to capture the implied autocorrelation structure, or it does but fails in our calibration experiment as a side-effect of the models' inability to reproduce the term structure of (a)symmetry. We believe the first explanation is most likely, though both could hold simultaneously. The reason is that we in section 5 calibrate a comparable one-factor rough Bergomi type model jointly to SPX and VIX options on a date where the term structure of (a)symmetry is negligible and still conclude that the short and long time-scale properties are not adequately separated. It is, at a glance, surprising that the fractional kernel, at least for these models, supposedly lacks flexibility as it already has an apparently rich structure by inducing mean-reversion at multiple time scales; we refer to Abi Jaber and El Euch (2019b) and Abi Jaber (2019) for the latter.[†] The results of section 5 suggest that the volatility autocorrelation structure is better captured by a classical but two-factor volatility model or a quadratic rough Heston model.

Let us suggest another reason for the term structures of residual smile effect: the log-normal distribution and that

implied by rough Heston may not be heavy-tailed enough to reproduce the short term behaviour of volatility. On the former, the authors of Bennedsen *et al.* (2017a) e.g. show that log-volatility is better described by a normal inverse Gaussian (NIG) distribution with heavier tails than a normal distribution. It is likewise a common observation that Heston type models are too light-tailed, see e.g. Gatheral (2008). The results of section 5 will though show that, while there does appear to be a distributional problem in some market scenarios in fitting to SPX options for at least the log-normal distribution, most of the residual calibration error can be eliminated by using a two-factor but yet log-normal model. We therefore believe that the poor SPX fits are mainly explainable by some combination of an inflexible volatility autocorrelation structure and the lack of a term structure for the smile (a)symmetry. The distributional problem is more apparent when we consider VIX options where log-normal models tend to produce almost flat smiles in contrast with the upward sloping ones typically observed in the market; see e.g. Horvath *et al.* (2020).[§]

Although the term structures of correlation and smile effect are robust across the different volatility levels, there is evidence that at least some of the effects become more pronounced when volatility is low. This is especially true of $\tilde{\eta}(T)/\eta$ under rough Bergomi and also slightly of $\tilde{\rho}(T)/\rho$ under rough Heston. It is harder to spot any meaningful level dependence for $\tilde{\rho}(T)/\rho$ under rough Bergomi or $\tilde{\nu}(T)/\nu$ under rough Heston. However, in either case, the relative wRMSE improves significantly more for both models when volatility is low indicating that the problems are much more visible in those scenarios. This is reflected in figure 8 where the VIX index is at 17.98 on September 4, 2012, and much lower at 11.23 on January 13, 2017, where the fits are notably worse.

To summarise, we have found evidence that the fractional kernel as used in the rough Bergomi and rough Heston models is not flexible enough to separate the short and long time scale

of the term structure of smile effect when there is also a significant term structure of (a)symmetry.

[†] One can check that $t^{H-1/2} = \int_0^\infty e^{-\gamma t} \mu(d\gamma)$ where $\mu(d\gamma) = \gamma^{-H-1/2} \Gamma(1/2-H)^{-1} d\gamma$. As described in the cited papers, we may then express rough volatility as a superposition of an infinite set of semi-martingales that mean-revert at both very slow and fast exponential rates with weights set by μ . Typical one-factor classical models correspond to only a single exponential term. The Heston equation (4) can e.g. be rewritten as $V_t = \xi_0(t) + \int_0^t e^{-\kappa(t-s)} \eta \sqrt{V_s} dW_{2,s}$; use Ito's lemma on $X_t := e^{\kappa t} V_t$ and rewrite. However, even if the rough fractional kernel generates mean-reversion at multiple time-scales in the sense just described, the measure μ is controlled by the single parameter H which limits the flexibility.

[‡] Related to the previous footnote, note also that rough Bergomi, at least under certain assumptions, tends to Black-Scholes in the limit $H \rightarrow 0$ as shown in Forde *et al.* (2022). The limit $H \rightarrow 0$ is precisely that in which we can further decouple the smile effect at short and long expiries and the result therefore highlights again the limitations of the fractional kernel.

[§] To solve this problem, the authors of Horvath *et al.* (2020) introduce a class of modulated Volterra processes for volatility and show that the VIX smiles can be captured using such a model with stochastic volatility-of-volatility.

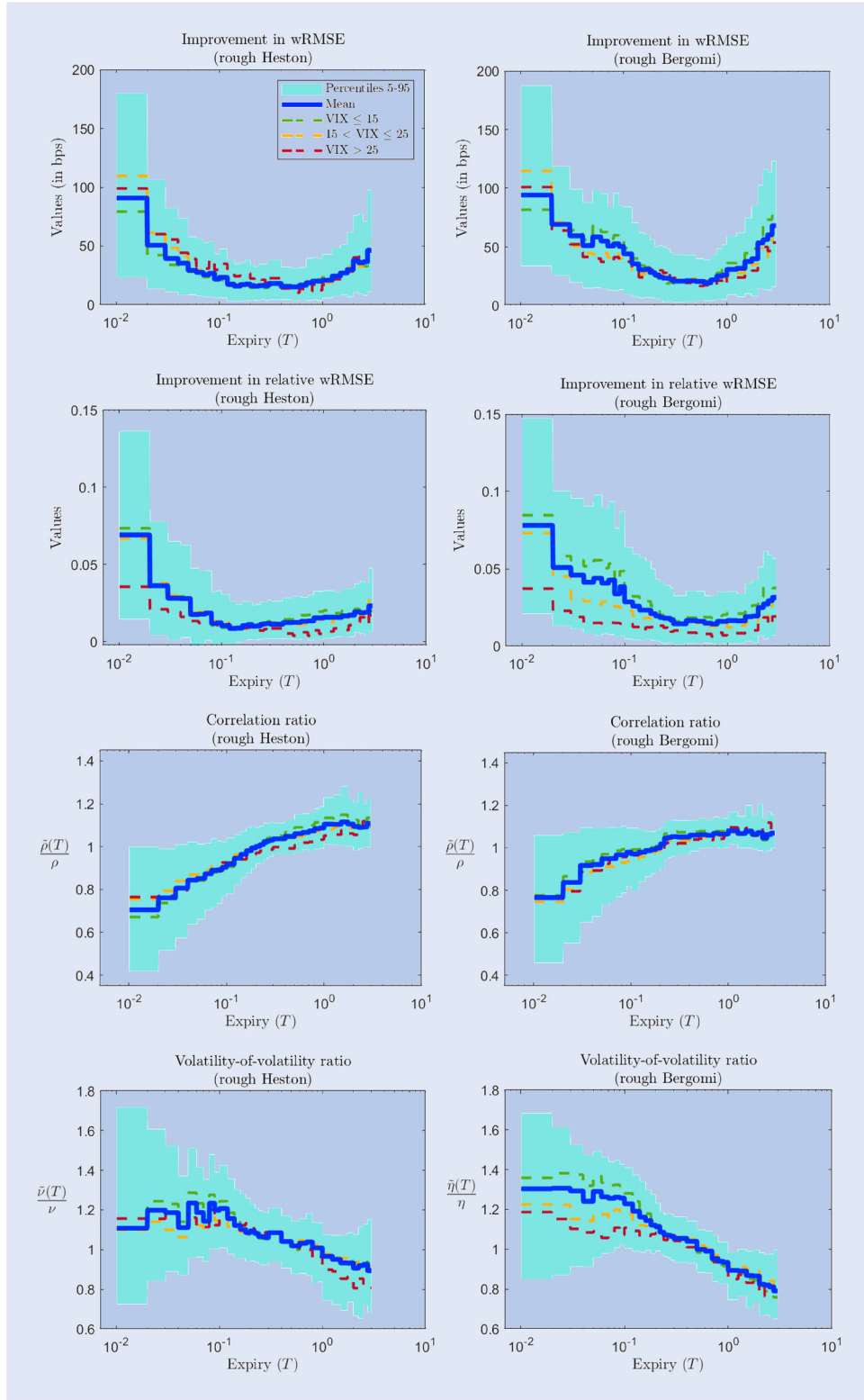


Figure 10. Expiry-by-expiry recalibrations. Shaded areas cover percentiles 5 to 95, solid lines show mean values, dashed lines show mean values categorised by the values of the VIX index.

properties of volatility that is implied by quoted SPX options. Furthermore, the term structure of smile (a)symmetry that we observe, suggests to model volatility two-factor where one factor should be less negatively correlated with the S&P 500 index and have relatively more influence on short expiries. The latter suggests that this factor should also be noisier and mean-revert faster.

5. Calibration of more advanced models

In search of a solution to the skew-curvature term structure problem and to examine the ability of rough volatility for the joint SPX-VIX calibration problem, we in this section calibrate a series of mostly more advanced models, including two-factor volatility models, to SPX options and jointly to

SPX and VIX options. However, with the increased model complexity, we find it is considerably more challenging to find reasonable moneyness bounds to train neural networks. Moreover, it is unclear how many extra training samples are needed when we add many more parameters to our models. In our experience, the time needed to compute the training datasets, although done ‘offline’, cannot be completely disregarded; recall that we spent around a week to compute the datasets for part one of our analysis. For these reasons, we therefore in this section perform the calibrations directly with Monte Carlo and only calibrate to three dates, though, they will reflect different market scenarios. We leave it for future work to see if neural networks can be implemented robustly and in a reasonable time for the more advanced models that we consider.[†]

5.1. Model selection

We assume yet that the S&P 500 index has risk-neutral dynamics of the form (1). For the volatility models of the present section, we introduce a third Brownian motion $(W_{3,t})_{t \geq 0}$ and now assume that $dW_{1,t}dW_{2,t} = \rho_{12}dt$, $dW_{1,t}dW_{3,t} = \rho_{13}dt$, $dW_{2,t}dW_{3,t} = \rho_{23}dt$, where $\rho_{12}, \rho_{13}, \rho_{23} \in [-1, 1]$ are so the associated correlation matrix is positive semi-definite. Moreover, we let $(\mathcal{F}_t)_{t \geq 0}$ be generated by $(W_{1,t}, W_{2,t}, W_{3,t})_{t \geq 0}$ and augmented. We will not always use all three Brownian motions.

All volatility models except one will be driven by Gaussian processes of the form

$$Y_t = \int_0^t K(t-s) dW_s, \quad t \geq 0, \quad (19)$$

where W_t represents either $W_{2,t}$ or $W_{3,t}$ and $K \in L^2_{\text{loc}}(\mathbb{R}_+, \mathbb{R})$. In the most flexible case, we consider the so-called *gamma* kernel defined by

$$K(t) = e^{-\lambda t} t^\alpha, \quad t > 0, \lambda \geq 0, \alpha \in (-1/2, 0], \quad (20)$$

nesting the rough fractional ($\lambda = 0$) and exponential ($\alpha = 0$) kernels. The latter results in an Ornstein-Uhlenbeck process if used in (19). We consider the gamma kernel for two reasons: Firstly, with $\lambda > 0$, we get bounded asymptotic variance in that $\lim_{t \rightarrow \infty} \text{Var}(Y_t) = \int_0^\infty e^{-2\lambda s} s^{2\alpha} ds < \infty$.[‡] If $\lambda = 0$, we have $\text{Var}(Y_t) = \int_0^t s^{2\alpha} ds = \frac{1}{2\alpha+1} t^{2\alpha+1} \rightarrow \infty$ in the same limit. Asymptotically bounded variance is desirable for volatility which reasonably can be assumed stationary. Secondly, the kernel allows some separate control over the singular part (via α) and the long term behaviour (via λ).

In what follows, we outline four general volatility models. We consider then several ones nested therein. We provide an

overview of the nested models that we consider in table 5. For each we show an identifier (i.e. label) and specify its definition from a general model. Furthermore, we have introduced simplified parameter notation where we have deemed it more natural. For some of the models, we will restrict the parameters when we calibrate to SPX options. Also, some of them will not be calibrated to the joint problem. Nevertheless, we believe our model selection for both calibration problems is more than rich enough for a detailed and meaningful analysis.

The first general volatility model that we consider is a shifted mixture two-factor rough Bergomi type model defined by

$$V_t = \zeta_0(t) (\mu X_{1,t} + (1 - \mu) X_{2,t}) + c \quad (21)$$

$$X_{i,t} = \mathcal{E}(\eta_i \delta_i (\theta_i Y_{1,t} + (1 - \theta_i) Y_{2,t})) \quad (22)$$

$$Y_{i,t} = \int_0^t K_i(t-s) dW_{i+1,s}, \quad i = 1, 2, t \geq 0, \quad (23)$$

where $\zeta_0 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is an appropriate deterministic function, $K_i(t) = e^{-\lambda_i t} t^{\alpha_i}$, $t > 0$, $\lambda_i \geq 0$, $\alpha_i \in (-\frac{1}{2}, 0]$, $i = 1, 2$, $\eta_1, \eta_2, c \geq 0$, $\mu, \theta_1, \theta_2 \in [0, 1]$. Note that $\xi_0(t) = \zeta_0(t) + c$ for $t \geq 0$. We choose the constants δ_1 and δ_2 so η_1 and η_2 represent normalised volatility parameters for $X_{1,t}$ and $X_{2,t}$. To explain how, define first

$$\begin{aligned} g(t; \theta, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \rho_{23}) \\ := \theta^2 \int_0^t K_1(s)^2 ds + (1 - \theta)^2 \int_0^t K_2(s)^2 ds \\ + 2\rho_{23}\theta(1 - \theta) \int_0^t K_1(s)K_2(s) ds, \\ t \geq 0, \theta \in [0, 1]. \end{aligned} \quad (24)$$

Let then $i \in \{1, 2\}$ and recall for the next part that

$$\text{Var}(Z_1 + Z_2) = \text{Var}(Z_1) + \text{Var}(Z_2) + 2\text{Cov}(Z_1, Z_2) \quad (26)$$

for sufficiently integrable random variables Z_1 and Z_2 . Note by the Ito isometry that

$$\text{Var}(\log(X_{i,t})) = (\eta_i \delta_i)^2 g(t; \theta_i, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \rho_{23}).$$

We then set

$$\delta_i := g(1; \theta_i, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \rho_{23})^{-\frac{1}{2}} \quad (27)$$

such that $\text{Var}(\log(X_{i,1})) = \eta_i^2$ always. It is worth noting that g can be written on semi-explicit form: Let $i, j \in \{1, 2\}$ and assume that $\lambda_i \neq 0$ or $\lambda_j \neq 0$. Then

$$\begin{aligned} \int_0^t K_i(s)K_j(s) ds &= \int_0^t e^{-(\lambda_i + \lambda_j)s} s^{\alpha_i + \alpha_j} ds \\ &= (\lambda_i + \lambda_j)^{-\alpha_i - \alpha_j - 1} \gamma(\alpha_i + \alpha_j + 1, \\ &\quad (\lambda_i + \lambda_j)t) \end{aligned}$$

where $\gamma(s, x) = \int_0^x u^{s-1} e^{-u} du$ denotes the lower incomplete gamma function. If $\lambda_i = \lambda_j = 0$ we get $\int_0^t K_i(s)K_j(s) ds =$

[†] We definitely think that the application of neural networks to the more advanced models should be explored. The Monte Carlo calibrations of the present section each took at least 10 minutes running on a server. In the absence of faster computational methods (Fourier pricing techniques, expansion formulas, etc.) or other calibration tricks (e.g. smart initial guesses), neural networks therefore still seem like the most promising and generally applicable tool for making calibration practically viable.

[‡] The singular part is integrable and the non-singular part is, up to scale, bounded by $\int_0^\infty e^{-2\lambda s} ds = \frac{1}{2\lambda}$.

Table 5. Model specifications.

Model identifier	Equations	Specification
S-RB-F	(21)–(23)	$\mu = 1, \theta_1 = 1, \eta := \eta_1, \rho := \rho_{12}, \lambda_1 = 0, \alpha := \alpha_1$
S-RB	(21)–(23)	$\mu = 1, \theta_1 = 1, \eta := \eta_1, \rho := \rho_{12}, \lambda := \lambda_1, \alpha := \alpha_1$
S-2F-B	(21)–(23)	$\mu = 1, \theta := \theta_1, \eta := \eta_1, \alpha_1 = \alpha_2 = 0$
S-2F-RB	(21)–(23)	$\mu = 1, \theta := \theta_1, \eta := \eta_1$
S-M-1F-RB	(21)–(23)	$\theta_1 = \theta_2 = 1, \rho := \rho_{12}, \lambda := \lambda_1, \alpha := \alpha_1$
S-M-2F-RB	(21)–(23)	No parameter restrictions or simplified notation
S-M-2F-RHyp	(28)–(30)	$\zeta_{2,0}(t) = \zeta_{1,0}(t) + \epsilon$ where $\epsilon \in \mathbb{R}$
QRH	(43)–(44)	$\eta = 1, \rho := \rho_{12}$
S-M-2F-QHyp	(45)–(48)	$\zeta_{2,0}(t) = \zeta_{1,0}(t) + \epsilon$ where $\epsilon \in \mathbb{R}, \alpha_1 = 0, \alpha_2 = 0$
S-M-2F-QRHyp	(45)–(48)	$\zeta_{2,0}(t) = \zeta_{1,0}(t) + \epsilon$ where $\epsilon \in \mathbb{R}$

* With the S-prefix removed we refer to the same model with $c = 0$.

$(\alpha_i + \alpha_j + 1)^{-1} t^{\alpha_i + \alpha_j + 1}$. Plugging into (24)–(25) we obtain an expression for g .

For Monte Carlo pricing under (21)–(23), we simulate $Y_{1,t}$ and $Y_{2,t}$ with the hybrid multifactor scheme of Rømer (2022). Plugging into (21)–(22), we obtain then values of V_t . We compute the VIX index as described in Rømer (2022) (up to adding the shift c) and simulate S_t with a log-Euler scheme. For the other models of this section, we likewise use a log-Euler scheme for S_t .

We justify the generality of the model (21)–(23) as follows: We have included two stochastic factors as the results of section 4 suggest. Furthermore, to create a skew for VIX options we allow $\mu \in (0, 1)$ (i.e. a mixture distribution); the works (Bergomi 2008, De Marco 2018, Guyon 2018) also use this idea. As illustrated in figure 11, a simple shift ($c > 0$) can though also generate a skew and we therefore include both features for most flexibility.[†] The Heston++ model of Pacati *et al.* (2014) and Pacati *et al.* (2018) with success also uses a volatility shift, though time-dependent.

We will test six models nested in (21)–(23), that is, up to restricting $c = 0$ or not; we refer to rows 1–6 of table 5. The idea behind our labelling is as follows: We write ‘S’ for a shifted model ($c \geq 0$ is unrestricted), ‘M’ for a mixture model ($\mu \in [0, 1]$ is unrestricted), ‘RB’ for a rough Bergomi type model, ‘B’ for a non-rough Bergomi type model. Also, by ‘1F’ and ‘2F’, we refer to, respectively, one- and two-factor volatility. For the rough models, our labelling implicitly assumes that the (rough) gamma kernel (20) is used. That is, unless the letter ‘F’ is used, in which case we refer to the pure fractional kernel ($\lambda = 0$). A careful inspection of the labels of table 5 show that we do not always use all relevant letters. This is to limit the length of the names. We encourage the reader to become familiar with the model definitions, labels, and the simplified parameter notation as it will greatly enhance the reading of the rest of the paper.

For the models nested in (21)–(23), we fix $c = 0$ when we calibrate to only SPX options so as to avoid overfitting the shift parameter to that market. We then drop the S-prefix in the labels. Also, to simplify our analysis, we do not calibrate to the joint problem the one-factor models S-RB-F and S-RB or the shifted two-factor Bergomi model S-2F-B. These models, however, are nested in the remaining ones, which include

a shifted two-factor rough Bergomi type model (S-2F-RB) and shifted one- and two-factor mixture rough Bergomi type models (S-M-1F-RB and S-M-2F-RB).

When it comes to the models nested in (21)–(23) with mixture terms, i.e. S-M-1F-RB, S-M-2F-RB, and their $c = 0$ restrictions, some of our calibrations result in distributions of V_t that are too heavy tailed for accurate price estimation. To get around this problem, we borrow an idea from Kahl and Jäckel (2006) by considering the hyperbolic transformation $f_{\text{hyp}}(x) := x + \sqrt{x^2 + 1}$ as an alternative to the exponential transformation $f_{\text{exp}}(x) := e^x$ used in (21)–(23). The two functions are similar in that they both are monotone mappings from \mathbb{R} onto \mathbb{R}_+ and behave alike near the origin; see their equations (41), (42), and figure 1. Importantly, however, the hyperbolic transformation grows much slower at large positive values and also decays slower at very negative values. Because of this, the authors’ hyperbolic model for $\sigma_t := \sqrt{V_t}$, and which is defined as $\sigma_t = \sigma_0 f_{\text{hyp}}(Y_t)$ where Y_t is an Ornstein-Uhlenbeck process, has thinner tails than the otherwise comparable log-normal model $\sigma_t = \sigma_0 f_{\text{exp}}(Y_t)$ proposed in Scott (1987); see e.g. their figure 2.

In our context, a natural model based on the hyperbolic transformation is

$$V_t = \mu f_{\text{hyp}}(X_{1,t}) + (1 - \mu) f_{\text{hyp}}(X_{2,t}) + c \quad (28)$$

$$X_{i,t} = \eta_i \delta_i (\theta_i Y_{1,t} + (1 - \theta_i) Y_{2,t}) \quad (29)$$

$$Y_{i,t} = \zeta_{i,0}(t) + \int_0^t K_i(t-s) dW_{i+1,s}, \quad i = 1, 2, t \geq 0, \quad (30)$$

where $\mu, c, \eta_1, \eta_2, \theta_1, \theta_2, K_1, K_2$ are as for the model (21)–(23), the constants (δ_1, δ_2) yet defined by (27), and $\zeta_{i,0} : \mathbb{R}_+ \rightarrow \mathbb{R}$, $i = 1, 2$, are appropriate deterministic functions. There are no problems with the moments of V_t since $f_{\text{hyp}}(x)$ grows linearly in the limit $x \rightarrow \infty$ so those of V_t are bounded by appropriate moments of the Gaussians $X_{1,t}$ and $X_{2,t}$; the same logic applies to the *quadratic* hyperbolic model that we define on pages 22–23. Let now

$$\begin{aligned} \zeta_{i,t}(u) &:= \zeta_{i,0}(u) + \int_0^t K_i(u-s) dW_{i+1,s}, \\ i &= 1, 2, 0 \leq t \leq u, \end{aligned} \quad (31)$$

[†] Large shifts relative to the volatility level can though flatten the smiles for options on S ; c.f. the top right plot.

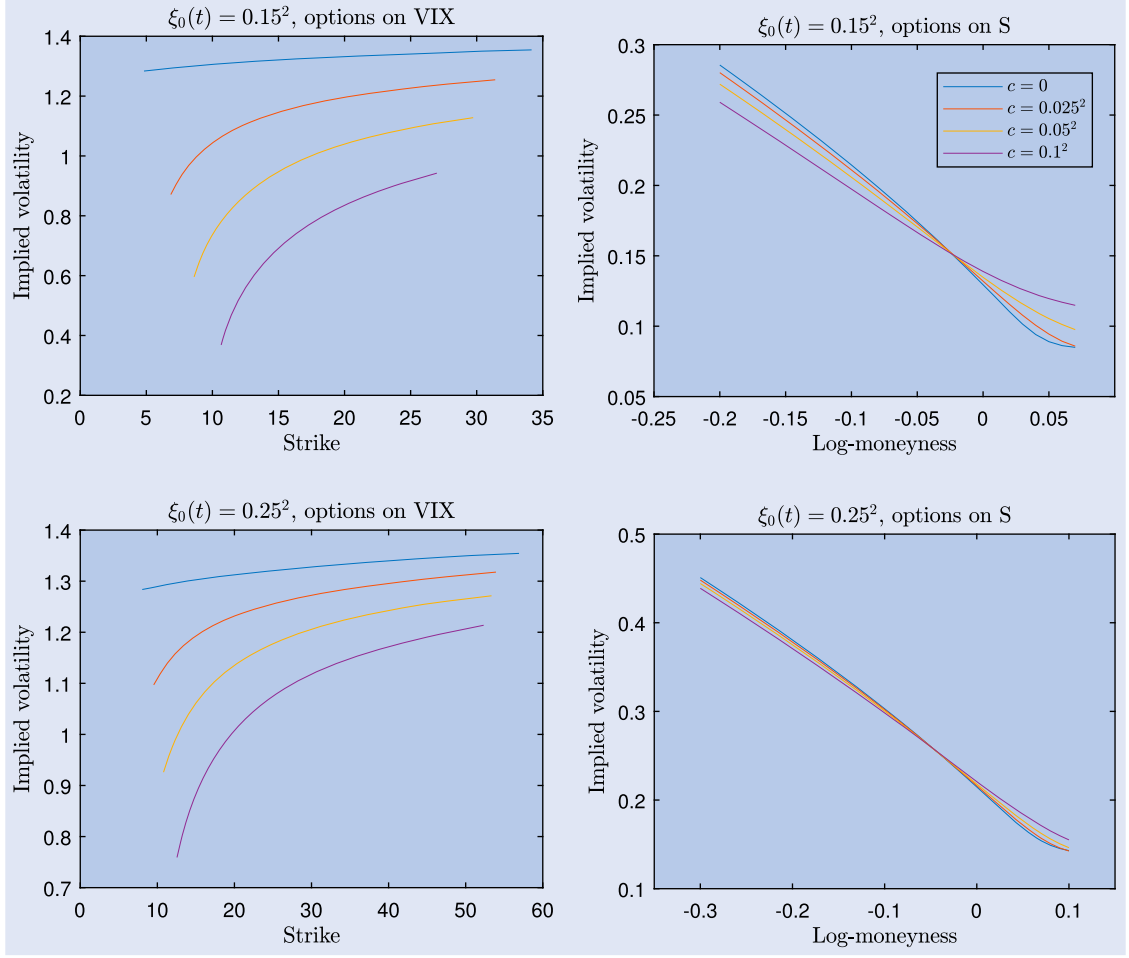


Figure 11. Option smiles at the expiry 0.1 for the shifted rough Bergomi model (S-RB-F) defined in table 5 with $\alpha = -0.4$, $\eta = 2.1$, $\rho = -0.9$, shown for different values of ξ_0 , which is assumed flat, and c ; recall that $\xi_0(t) = \zeta_0(t) + c$.

and note that

$$Y_{i,u} = \zeta_{i,t}(u) + \int_t^u K_i(u-s) dW_{i+1,s}, \quad i = 1, 2, \quad 0 \leq t \leq u, \quad (32)$$

where the first term is measurable with respect to \mathcal{F}_t , the second term independent.

We briefly justify the time-dependence in $\zeta_{1,0}$ and $\zeta_{2,0}$: Let $i \in \{1, 2\}$ and define for $t, t_0 \geq 0$: $Y_{i,t}^{t_0} := Y_{i,t_0+t}$, $W_{i+1,t}^{t_0} := W_{i+1,t_0+t} - W_{i+1,t_0}$, $\mathcal{F}_t^{t_0} := \mathcal{F}_{t_0+t}$. Fix $t_0 \geq 0$. By (32) then

$$\begin{aligned} Y_{i,t}^{t_0} &= \zeta_{i,t_0}(t_0 + t) + \int_{t_0}^{t_0+t} K_i(t_0 + t - s) dW_{i+1,s} \\ &= \zeta_{i,t_0}(t_0 + t) + \int_0^t K_i(t - s) dW_{i+1,s}^{t_0}, \quad t \geq 0. \end{aligned}$$

Since $(W_{i+1,t}^{t_0})_{t \geq 0}$ is Brownian motion with respect to $(\mathcal{F}_t^{t_0})_{t \geq 0}$ and ζ_{i,t_0} is \mathcal{F}_{t_0} -measurable, the distribution of $(Y_{i,t}^{t_0})_{t \geq 0}$ conditionally on \mathcal{F}_{t_0} is equivalent to that at time zero but with an updated initial term structure $t \mapsto \zeta_{i,t_0}(t_0 + t)$. In particular, even if $\zeta_{i,t_0}(t_0 + \cdot)$ is flat at $t_0 = 0$, it generally is not for $t_0 > 0$; note (31). Heuristically, we can therefore think of any time-dependence in $\zeta_{i,0}$ as an expression of the historical path-dependence that has occurred before time zero. It also means that movements in the curves $\zeta_{1,t_0}(t_0 + \cdot)$ and $\zeta_{2,t_0}(t_0 + \cdot)$ in theory can be hedged under the model.

REMARK 1 If we apply the same logic to (21)–(23) we can justify both a time-dependent initial forward variance curve ξ_0 , equivalently ζ_0 , and a time-dependent mixing parameter μ . That we keep μ constant in the definition is a simplification to avoid overfitting to the time-dependency. Similar arguments will justify any time-dependencies in the remaining models that will be defined.

We consider a single model nested in (28)–(30) for our calibrations; we refer to the label S-M-2F-RHyp in table 5. To avoid overfitting to the time-dependency, we have restricted $\zeta_{2,0}(t) = \zeta_{1,0}(t) + \epsilon$ for a number $\epsilon \in \mathbb{R}$. We again fix $c = 0$ when we calibrate to SPX options only.

For (28)–(30), we likewise simulate $Y_{1,t}$ and $Y_{2,t}$ with the hybrid multifactor scheme. Plugging into (28)–(29), we obtain then the instantaneous variance and may thereafter simulate S_t with a log-Euler scheme. Computing the VIX index as part of the simulations is however less trivial and not covered by the discussions in Rømer (2022). We therefore explain our approach in the following.

We start with a trapezoidal rule applied to (18) as

$$\text{VIX}_t^2 \approx \frac{100^2}{n_v} \sum_{i=0}^{n_v} a_i \xi_t(t + \tau_i), \quad t \geq 0, \quad (33)$$

with $n_v \geq 2$ an integer and $a_0 = a_{n_v} = \frac{1}{2}$, $a_i = 1$, $i = 1, \dots, n_v - 1$, $\tau_i = \frac{i}{n_v} \Delta$, $i = 0, \dots, n_v$. To compute VIX_t as part of the simulations it then suffices to compute $\xi_t(u)$ for specific u 's. Note that

$$\xi_t(u) = \mu E_t(f_{\text{hyp}}(X_{1,u})) + (1 - \mu) E_t(f_{\text{hyp}}(X_{2,u})) + c, \quad 0 \leq t \leq u. \quad (34)$$

Define

$$\chi_{i,t}(u) := \eta_i \delta_i (\theta_i \zeta_{1,t}(u) + (1 - \theta_i) \zeta_{2,t}(u)) \quad (35)$$

$$v_i(u - t) := \eta_i \delta_i \sqrt{g(u - t; \theta_i, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \rho_{23})}, \quad i = 1, 2, \quad 0 \leq t \leq u. \quad (36)$$

By (29) and (32), we for $i \in \{1, 2\}$ and (t, u) such that $0 \leq t \leq u$ have

$$E_t(f_{\text{hyp}}(X_{i,u})) = E_t(f_{\text{hyp}}(\eta_i \delta_i \{\theta_i Y_{1,u} + (1 - \theta_i) Y_{2,u}\})) \quad (37)$$

$$= E_t\left(f_{\text{hyp}}\left(\eta_i \delta_i \left\{\theta_i \zeta_{1,t}(u) + (1 - \theta_i) \zeta_{2,t}(u)\right\}\right)\right) \quad (38)$$

$$+ \eta_i \delta_i \left\{ \theta_i \int_t^u K_1(u - s) dW_{2,s} \right. \quad (39)$$

$$\left. + (1 - \theta_i) \int_t^u K_2(u - s) dW_{3,s} \right\} \Bigg). \quad (40)$$

Recall that $\zeta_{1,t}(u)$ and $\zeta_{2,t}(u)$ are \mathcal{F}_t -measurable and that the stochastic integrals of (39)–(40) are independent of \mathcal{F}_t and therefore are conditionally normally distributed as they are unconditionally. By this, (26), the Ito isometry, (35)–(36), we conclude that

$$E_t(f_{\text{hyp}}(X_{i,u})) = E(f_{\text{hyp}}(x + v_i(u - t)Z)) \Big|_{x=\chi_{i,t}(u)}, \quad i = 1, 2, \quad 0 \leq t \leq u, \quad (41)$$

where $Z \sim \mathcal{N}(0, 1)$. Given knowledge of the terms in (35)–(36), the problem then reduces to that of computing unconditional expectations of a transformed standard normal variable. Note that values of $v_i(u - t)$ are readily computable via (36) and the previously stated expressions for g . As the variables are deterministic, the cost of evaluating them is small when spread across many paths. Even better as $v_i(u - t)$ only depends on the time-to-maturity $u - t$, we can reuse them when computing VIX_t at multiple time points t . In contrast, values of $\chi_{i,t}(u)$ require knowledge of $\zeta_{1,t}(u)$ and $\zeta_{2,t}(u)$ which depend on \mathcal{F}_t . Fortunately, however, these terms are relatively easy to compute under the hybrid multifactor scheme; we refer to Rømer (2022) for the details.

Let us then discuss how to evaluate (41) given knowledge of (35)–(36). Define

$$h(x, y) := \int_{\mathbb{R}} f_{\text{hyp}}(x + yz) \phi(z) dz, \quad (x, y) \in \mathbb{R}^2, \quad (42)$$

where $\phi(z) := (2\pi)^{-\frac{1}{2}} e^{-z^2/2}$, $z \in \mathbb{R}$. Then $E_t(f_{\text{hyp}}(X_{i,u})) = h(\chi_{i,t}(u), v_i(u - t))$ for $i = 1, 2$, and $0 \leq t \leq u$, which

reduces the problem to evaluating h . To our knowledge, no analytical expression exists. We therefore use numerical integration with Matlab's *integral* function. However, evaluating $h(\chi_{i,t}(u), v_i(u - t))$, $i = 1, 2$, with numerical integration across multiple time points t , forward variance maturities u , and sample paths, quickly becomes very time consuming. For our implementation, we therefore, given a set of simulated paths, for each $i \in \{1, 2\}$ and relevant time-to-maturity $u - t$, compute $h(\cdot, v_i(u - t))$ on a smaller grid of points that cover the necessary range and then use linear interpolation. We find that this can greatly speed up the computations.

Next, we consider a quadratic rough Heston model defined by

$$V_t = a(Z_t - b)^2 + c \quad (43)$$

$$Z_t = \zeta_0(t) + \int_0^t K(t - s) \eta \sqrt{V_s} dW_{2,s}, \quad t \geq 0, \quad (44)$$

where $a, c, \eta \geq 0$, $b \in \mathbb{R}$, $K(t) = e^{-\lambda t} t^\alpha$, $\lambda \geq 0$, $\alpha \in (-\frac{1}{2}, 0]$, and $\zeta_0 : \mathbb{R}_+ \rightarrow \mathbb{R}$ is locally Hölder continuous of all orders strictly less than $\alpha + \frac{1}{2}$. A unique continuous strong solution exists by Abi Jaber and El Euch (2019a, Theorem A.1).[†] Define $f^{\alpha, \lambda}(t) := \lambda t^{\alpha-1} E_{\alpha, \alpha}(-\lambda t^\alpha)$, $t > 0$, $\alpha \in (\frac{1}{2}, 1)$, $\lambda > 0$, where

$$E_{\alpha, \beta}(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + \beta)}, \quad z, \alpha, \beta \in \mathbb{C},$$

$$\text{Re}(\alpha), \text{Re}(\beta) > 0,$$

is the so-called *Mittag-Leffler* function. The original model of Gatheral *et al.* (2020) then fits into the formulation (43)–(44) with $K(t) = f^{\alpha, \lambda}(t)$, $\zeta_0(t) = \int_0^t f^{\alpha, \lambda}(t - s) \theta_0(s) ds$, $t > 0$, where θ_0 is a deterministic function. We believe our version has comparable generality as ζ_0 in both cases control the term structure of $(Z_t)_{t \geq 0}$ (directly or via θ_0) and since the α 's (which are equivalent up to a shift) both control the singular part of K while the λ 's control the decay at larger values. We use the hybrid multifactor scheme of Rømer (2022) to simulate the model. To compute the VIX index, we solve equation (105) of that paper, also with the hybrid multifactor scheme.[‡] For normalisation we set $\eta = 1$ as suggested in Gatheral *et al.* (2020) for the original model. The version we calibrate is stated in table 5 with the label QRH. In keeping somewhat with the original formulation, we allow $c \geq 0$ also when we calibrate to SPX options only.

Inspired by (28)–(30) and (43)–(44), we consider lastly a two-factor model that uses both the quadratic and hyperbolic transformations. Specifically, we consider the model

$$V_t = \mu X_{1,t} + (1 - \mu) X_{2,t} + c \quad (45)$$

[†] To use the theorem we must check: (1) That K satisfies hypothesis (H_0) of the paper with their $\gamma = 2\alpha + 1$. This follows from parts (ii) and (iv) of Abi Jaber *et al.* (2019, example 2.3). (2) That the diffusion coefficient of (44) is Lipschitz continuous in the value of Z_s . We here refer to Rømer (2022).

[‡] A more efficient computational method for the VIX index is devised in also Rømer (2022) which involves discretising equation (109) of the paper. However, our quadratic rough Heston implementation was written at the time of an earlier version of Rømer (2022) where this equation did not appear. This is why we compute the VIX index as described instead.

$$X_{i,t} = f_{\text{hyp}}(Z_{i,t}^2 - d_i) - f_{\text{hyp}}(-d_i), \quad (46)$$

$$Z_{i,t} = \eta_i \delta_i (\theta_i Y_{1,t} + (1 - \theta_i) Y_{2,t}), \quad (47)$$

$$Y_{i,t} = \zeta_{i,0}(t) + \int_0^t K_i(t-s) dW_{i+1,s}, \quad i = 1, 2, t \geq 0, \quad (48)$$

where $\mu, \theta_1, \theta_2 \in [0, 1]$, $\eta_1, \eta_2, c \geq 0$, $d_1, d_2 \in \mathbb{R}$, (δ_1, δ_2) is defined by (27), and we for $i \in \{1, 2\}$ have $K_i(t) = e^{-\lambda_i t} t^{\alpha_i}$, $t > 0$, $\lambda_i \geq 0$, $\alpha_i \in (-\frac{1}{2}, 0]$, with $\zeta_{i,0} : \mathbb{R}_+ \rightarrow \mathbb{R}$ an appropriate deterministic function. The parameters d_1 and d_2 are introduced to allow control over the minima of the terms $Z_{i,t}^2 - d_i$, $i = 1, 2$. The processes $X_{1,t}$ and $X_{2,t}$ are shifted so the minimum value they can attain is zero. We simulate the model akin to (28)–(30) and likewise use the approach outlined on pages 21–22 for computing the VIX index—it carries over with only trivial changes. We calibrate a version with gamma kernels and one with pure exponential kernels; they are defined with labels S-M-2F-QRHyp and S-M-2F-QHyp in table 5. We again restrict $\zeta_{2,0}(t) = \zeta_{1,0}(t) + \epsilon$ for a number $\epsilon \in \mathbb{R}$. We also fix $c = 0$ when we calibrate to SPX options only.

REMARK 2 As noted, we use the hybrid multifactor scheme of Rømer (2022) to simulate the models of this section. There are several reasons for this: Firstly, the scheme convergences reasonably fast for the Gaussian case and for the quadratic rough Heston equation; evidence from the paper suggest that this is also true for Hurst exponents reasonably close to zero. Secondly, the scheme is, at least slightly, faster than the hybrid scheme of Bennedsen *et al.* (2017b) that we used in part one of our analysis and facilitates easy computation of so-called *forward values* such as (31) and which are needed to compute the VIX index for our models.† We refer to Rømer (2022) for more details.

5.2. Calibration results

We calibrate to the dates October 14, 2011, September 4, 2012, and January 13, 2017; the last two were also considered in figure 8. The VIX index is respectively at 28.24, 17.98, and 11.23, and we therefore cover the case of high, typical, and low volatility. Our VIX options dataset is, like our SPX dataset, obtained from <https://datashop.cboe.com>. We filter both datasets as we filtered the SPX dataset in section 4 (in-the-money options are e.g. removed) except we do not filter the range of moneyness but instead the set of expiries. To select expiries, we consider the intervals defined by the time points $\{0, 0.01, 0.02, 0.03, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 1.5, 2, 2.5, 3\}$ and in each select the expiry, if any, closest to the midpoint. This gives respectively 10, 12, and 13 SPX expiries and 6, 5, and 7 VIX expiries for each of the dates. To compute Black-Scholes implied volatility for VIX options, we use the same expiry VIX futures as the underlying. We extract the futures prices from the observed VIX option put-call parities.

For the joint problem, we find that it works better to measure the error in terms of deviations from the bid-ask bounds. We therefore change to such an error measure: Consider a

given date and let N_s and N_v denote the number of observed SPX options, respectively, VIX options. Let $\sigma_{s,i}^{\text{bid}}, \sigma_{s,i}^{\text{ask}}, \sigma_{s,i}^{\text{model}}$, respectively, $\sigma_{v,i}^{\text{bid}}, \sigma_{v,i}^{\text{ask}}, \sigma_{v,i}^{\text{model}}$, denote the bid, ask and model implied volatility for the i 'th observed SPX option, respectively, VIX option. We then measure the calibration error by the number

$$\sqrt{\theta \sum_{i=1}^{N_s} w_{s,i} F(\sigma_{s,i}^{\text{bid}}, \sigma_{s,i}^{\text{ask}}, \sigma_{s,i}^{\text{model}})^2 + (1 - \theta) \sum_{i=1}^{N_v} w_{v,i} F(\sigma_{v,i}^{\text{bid}}, \sigma_{v,i}^{\text{ask}}, \sigma_{v,i}^{\text{model}})^2} \quad (49)$$

where $\theta \in [0, 1]$ and we for a set of numbers x, y and z so $x < y$ have defined

$$F(x, y, z) := \begin{cases} x - z & z < x \\ z - y & z > y \\ 0 & \text{otherwise} \end{cases}.$$

The terms $\{w_{s,i}\}_{i=1}^{N_s}, \{w_{v,i}\}_{i=1}^{N_v}$ are weights which we choose so each expiry carries the same total amount but otherwise normalise them so $\sum_{i=1}^{N_s} w_{s,i} = \sum_{i=1}^{N_v} w_{v,i} = 1$. Stated in words, we measure the error as the square root of a weighted average of squared deviations of the model implied volatilities from the bid-ask bounds. The θ -parameter allows control of the relative weight of each market.

When we calibrate to SPX options, we set $\theta = 1$. We then choose three initial guesses and minimise (49) under each; we here assume that $\zeta_0, \zeta_{1,0}$, and $\zeta_{2,0}$, whichever relevant, are flat. We use that of the solutions with the lowest error as an initial guess for a final calibration whose results we present and for which we assume that the relevant curves are piecewise constant between the observed SPX expiries.‡ For the joint problem, we set $\theta = 0.9$ which we find leads to a reasonable balance between the two markets. We then run a single optimisation with the solution from the final SPX calibration as an initial guess. The final calibrations are run with 50 000 paths.

We present the calibration results in tables 6, 7, and 9. Fits are shown in figures 12–16. For brevity and visibility only a small selection of the expiries and models are included in the plots. However, since the chosen expiries include the shortest and longest ones on each date, and because the models that are included more or less represent the different error levels that can be found in the tables, it will suffice for our discussion. The reader may also note that we do not report calibration errors for some of the log-normal mixture type models. The reason is, as alluded to earlier, that it sometimes is difficult to estimate prices with sufficient accuracy under these models. We discuss the issue in detail below. Only thereafter do we analyse the calibration results.

To illustrate the estimation problem, we examine under the calibrated models Monte Carlo estimation of the 1-year fair variance swap strike, i.e. the value of $K_{\text{vs}} := E(\frac{1}{T} \int_0^T V_t dt)$

† It is due to the timing of the papers that we have used the scheme of Bennedsen *et al.* (2017b) for part one of our analysis.

‡ Although we need ζ_0 locally Hölder *continuous* of all orders less than $\alpha + \frac{1}{2}$ for a solution to (44), we find no meaningful price differences when we replace the calibrated ζ_0 curves with approximating C^1 curves.

Table 6. Calibration to SPX options.

RB-F	η	ρ	α	$\bar{\xi}_0$						Error					
2011-10-14	2.00	-1.00	-0.43	0.0845						35					
2012-09-04	1.90	-0.80	-0.41	0.0347						41					
2017-01-13	2.63	-1.00	-0.45	0.0145						86					
RB	η	ρ	λ	α	$\bar{\xi}_0$						Error				
2011-10-14	1.85	-0.88	0.88	-0.36	0.0848						32				
2012-09-04	1.84	-0.75	1.00	-0.33	0.0354						36				
2017-01-13	2.26	-0.65	14.79	-0.12	0.0127						87				
2F-B	θ	η	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	$\bar{\xi}_0$			Error				
2011-10-14	0.90	2.03	-0.50	-0.96	0.27	71.73	1.17	0.0803			7				
2012-09-04	0.90	2.03	-0.35	-0.93	-0.02	92.81	1.46	0.0330			11				
2017-01-13	0.88	2.34	-0.29	-0.97	0.05	74.62	2.89	0.0133			39				
2F-RB	θ	η	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	$\bar{\xi}_0$	Error				
2011-10-14	0.77	2.04	-0.48	-0.97	0.25	50.50	0.96	-0.17	-0.06	0.0807	7				
2012-09-04	0.67	2.00	-0.27	-1.00	0.29	45.67	0.90	-0.27	-0.17	0.0337	11				
2017-01-13	0.91	2.38	-0.22	-0.97	-0.03	90.56	2.41	-0.01	-0.11	0.0134	38				
M-1F-RB	μ	η_1	η_2	ρ	λ	α	$\bar{\xi}_0$					Error			
2011-10-14	0.82	1.61	7.17	-0.84	1.18	-0.24	0.0897					*			
2012-09-04	0.65	1.58	9.98	-0.75	2.08	-0.05	0.0430					*			
2017-01-13	0.22	1.99	7.58	-1.00	0.46	-0.43	0.0496					*			
M-2F-RB	μ	θ_1	θ_2	η_1	η_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	$\bar{\xi}_0$	Error	
2011-10-14	0.22	0.27	0.80	0.87	2.37	-0.21	-0.99	0.23	88.78	0.79	-0.21	-0.27	0.0817	3	
2012-09-04	0.69	0.80	0.90	1.37	4.94	-0.27	-0.90	0.00	82.03	1.42	-0.06	-0.12	0.0402	*	
2017-01-13	0.18	0.00	0.08	1.55	8.11	-0.99	-0.81	0.85	181.10	3.26	-0.48	-0.09	0.0573	*	
QRH	a	b	c	ρ	λ	α	$\bar{\xi}_0$								Error
2011-10-14	0.09	0.78	0.0120	1.00	0.23	-0.43	0.1458								29
2012-09-04	0.11	0.42	0.0115	1.00	0.83	-0.42	0.1668								30
2017-01-13	0.21	0.08	0.0024	1.00	8.82	-0.42	-0.0029								70

Notes: Bar-notation denotes average values till 1 month. Errors are reported in basis points, though, some are left out where price estimation is deemed too difficult; we refer to the main text.

where $T = 1$. We focus on variance swaps as K_{vs} can be computed with other deterministic methods to compare with. Note that if we cannot estimate a variance swap with sufficient accuracy, we can expect similar problems when we price other types of derivatives with Monte Carlo. The results of our experiment are presented in table 8 where we show K_{vs} computed with deterministic methods (denoted \hat{K}_{vs}^{\det}) alongside statistics on a Monte Carlo estimator (denoted \hat{K}_{vs}^{mc}).

We have computed \hat{K}_{vs}^{\det} as follows: For the models nested in (21)–(23), we use the formula $K_{vs} = \frac{1}{T} \int_0^T \zeta_0(t) dt + c =: \hat{K}_{vs}^{\det}$, which we can (and will) evaluate without error as ζ_0 is assumed piecewise constant. For the models nested in (28)–(30) and (45)–(48), we discretise $\frac{1}{T} \int_0^T E(V_t) dt$ with a trapezoidal rule and compute values of $E(V_t)$ with numerical integration along the lines of pages 21–22. We use the same trapezoidal rule for quadratic rough Heston, except we compute $E(V_t)$ by solving equation (105) of Rømer (2022) with the hybrid multifactor scheme. To construct a single Monte Carlo sample \hat{K}_{vs}^{mc} , we simulate 50 000 paths (as for the calibrations) of $(V_t)_{t \in [0, T]}$, compute values of $\frac{1}{T} \int_0^T V_t dt$ with a trapezoidal rule, and then average the numbers. The

statistics in table 8 are based on 100 samples of \hat{K}_{vs}^{mc} . We use a fine discretisation with 2500 equidistant steps for the trapezoidal rules and to simulate $(V_t)_{t \in [0, T]}$. We can then in particular expect \hat{K}_{vs}^{\det} computed with *all* the deterministic methods to closely approximate the true values K_{vs} ; for the models nested in (21)–(23) we of course have equality regardless. Moreover, we can for \hat{K}_{vs}^{mc} expect a small distributional error and therefore that the Monte Carlo sampling error should dominate.

In terms of the results, note that the median estimates are all lower than the values computed with the deterministic methods. We believe this is a consequence of V_t being right-skewed which is natural given that it lies close to but is bounded below by zero. However, for some of the models with log-normal mixture terms, the differences are substantial. For example, S-M-2F-RB calibrated to the joint problem on September 4, 2012, has \hat{K}_{vs}^{\det} , which *equals* K_{vs} for this example, at 0.1385 whereas the median of \hat{K}_{vs}^{mc} is 0.0950. The 5–95 percentile interval is 0.0833–0.1506 which shows that \hat{K}_{vs}^{mc} also has a very wide distribution. We believe the problems are worse under the log-normal mixture type

Table 7. Calibration to SPX options (more results).

M-2F-RHyp	μ	θ_1	θ_2	η_1	η_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	$\bar{\xi}_{1,0}$	$\bar{\xi}_{2,0}$	Error		
2011-10-14	0.38	0.50	0.00	4.87	25.45	−1.00	−0.97	0.96	0.00	6.96	−0.44	−0.04	−4.07	−0.75	27		
2012-09-04	0.17	0.74	0.05	10.54	9.64	−0.99	−1.00	1.00	0.87	0.95	−0.44	0.00	−4.61	−2.34	15		
2017-01-13	0.94	0.01	0.71	47.80	17.66	−0.86	−0.67	0.19	2.07	0.54	−0.40	−0.01	−2.34	−3.09	65		
M-2F-QHyp	μ	θ_1	θ_2	η_1	η_2	d_1	d_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	$\bar{\xi}_{1,0}$	$\bar{\xi}_{2,0}$	Error		
2011-10-14	0.42	0.15	0.07	0.43	1.03	1.68	15.00	1.00	0.51	0.50	1.35	91.46	−0.70	−0.09	4		
2012-09-04	0.64	0.12	0.05	0.62	0.89	3.48	8.19	0.90	0.32	−0.01	1.99	232.77	−0.45	−0.02	12		
2017-01-13	0.02	0.09	0.06	1.30	0.78	8.13	6.78	0.99	0.72	0.65	1.45	108.85	−0.71	0.02	20		
M-2F-QRHyp	μ	θ_1	θ_2	η_1	η_2	d_1	d_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	$\bar{\xi}_{1,0}$	$\bar{\xi}_{2,0}$	Error
2011-10-14	0.97	0.08	0.18	1.21	0.80	21.95	0.80	1.00	0.52	0.45	1.58	129.11	0.00	−0.03	−0.53	−0.09	9
2012-09-04	0.19	0.27	0.49	1.00	0.48	8.43	4.12	0.91	0.26	−0.07	0.69	45.36	−0.16	−0.26	−1.42	−0.11	5
2017-01-13	0.93	0.02	0.17	0.98	0.96	11.79	5.07	0.99	0.59	0.47	2.09	133.36	0.00	−0.04	−0.50	0.00	22

models because the tails are heavier which should amplify the effects.

For evidence of the last point, we refer to tables 6 and 9, where we note that η_1 or η_2 typically are relatively large whenever we observe an estimation problem in table 8. For example, for the problematic model S-M-2F-RB calibrated to the joint problem on September 4, 2012, we have $\eta_1 = 2.00$ and $\eta_2 = 5.84$, whereas the maximum value of η for any of the non-mixture log-normal models, and for which no such estimation problems are observed, is $\eta = 3.03$.[†] We believe the reason that we do not observe the same estimation problems for the hyperbolic models despite also being mixtures, is that the tails decay faster under the hyperbolic transformation. This, of course, was precisely our motivation for introducing the hyperbolic transformation.

We turn our attention to the SPX calibrations and start with a discussion of the one-factor models. We note first that there is a very limited improvement, if any, when we extend rough Bergomi (RB-F) to that with the more general gamma kernel (RB); the calibration errors in fact only change from 35, 41, and 86 bps, to 32, 36, and 87 bps. The results are only slightly better under quadratic rough Heston (QRH) which result in the errors 29, 30, and 70 bps. The fits for RB-F and QRH are shown in figures 12 and 13. Similar to the results of section 4.3 both models fail, at least partially, to capture the term structure of smile (a)symmetry. The problem is barely noticeable on October 14, 2011, where volatility is high, but becomes increasingly visible on the other dates. We can report that the smiles for RB are almost indistinguishable from those of RB-F on October 14, 2011, and September 4, 2012, and that RB too fails to reproduce the term structure of smile (a)symmetry on January 13, 2017. That RB-F, RB, and QRH, calibrate

comparably indicates that the main issue with typical one-factor rough volatility models in calibrating to SPX options is to recreate the term structure of (a)symmetry. It is then somewhat meaningful that the fits only change little when we add more flexibility to the kernel (fractional vs. gamma) and compare different distributions (log-normal vs. quadratic rough Heston), though, see also the remark below.

REMARK 3 It is not entirely implausible that one could in theory generate a sufficiently pronounced (a)symmetry term structure with quadratic rough Heston. Consider this: By $V_t = a(Z_t - b)^2 + c$, the correlation between Z_t and V_t , and thus between S_t and V_t , changes sign depending on Z_t being above or below b .[‡] With the initial curve ζ_0 we can then choose a correlation term structure. The reason why the model does not capture the observed (a)symmetry term structure is likely because ζ_0 also controls the mean term structure, i.e. ξ_0 , and thus there is not enough flexibility overall.[§]

We consider now the two-factor models. To this end, we note first that the two-factor Bergomi model (2F-B) calibrates much better than any of the one-factor models with errors of 7, 11, and 39 bps.[¶] As we can observe in figures 12 and 13, the improvement is especially noticeable on the low volatility date. For the first volatility factor, we calibrate ρ_{12} in the range from -0.50 to -0.29 and λ_1 in the range from 71.73 to 92.81. For the second factor, we calibrate ρ_{13} in the range from -0.97 to -0.93 and λ_2 in the range from 1.17 to 2.89. Since we have $\theta \approx 0.90$ on all three dates, the first factor is much more volatile. The parameters are then in line with the hypothesis that we gave in section 4.3, where we suggested the need for two volatility factors, one of which should mean-revert faster, be more volatile, and less negatively correlated with the S&P 500 index. The version with the gamma kernel (2F-RB) resulted in essentially the same errors (7, 11, and 38 bps).

[†] We find that it generally is for $\eta > 3$ that it becomes difficult to estimate the mean of $Y := \exp(\eta X - \frac{1}{2}\eta^2)$ where $X \sim \mathcal{N}(0, 1)$. To illustrate the problem for the example of the main text note this: Let $\eta = 5.84$ and $N \in \mathbb{N}$, $\hat{\mu} := \frac{1}{N} \sum_{i=1}^N Y_i$, where Y_i , $i = 1, \dots, N$, are i.i.d. and distributed as Y . Sampling 10 000 values of $\hat{\mu}$ where $N = 50\,000$, we get a median of 0.0812 and a 5–95 percentile interval of 0.0257–1.0627. This should be compared to the true mean which is 1. Convergence is also very slow in N : Sampling 10 000 values of $\hat{\mu}$ where $N = 10^6$, we get a median of 0.2202 and a 5–95 percentile interval of 0.0985–1.4136.

[‡] A similar observation is made in Forde and Smith (2021) for the original version of the model.

[§] In Forde and Smith (2021) another potential drawback of the model relating the martingale property of S is discussed.

[¶] In Guyon (2022) there is also an example of the market at-the-money term structure of skew that is more consistent with a two-factor Bergomi model than with a power-law.

Table 8. Estimation of the 1-year fair variance swap strike K_{VS} with the Monte Carlo estimator \hat{K}_{VS}^{mc} .

Date and model	Calibration to SPX options				Joint SPX-VIX calibration			
	\hat{K}_{VS}^{det}	Median minus \hat{K}_{VS}^{det}	5th perc.	95th perc.	\hat{K}_{VS}^{det}	Median minus \hat{K}_{VS}^{det}	5th perc.	95th perc.
Oct. 14, 2011:								
RB-F	0.1086	−0.0011	0.1064	0.1085				
RB	0.1120	−0.0008	0.1099	0.1125				
2F-B	0.1167	−0.0018	0.1137	0.1162				
(S-)2F-RB	0.1152	−0.0018	0.1122	0.1150	0.1235	−0.0035	0.1186	0.1216
(S-)M-1F-RB	0.1345	−0.0209	0.1113	0.1263	0.1087	−0.0004	0.1071	0.1101
(S-)M-2F-RB	0.1164	−0.0026	0.1124	0.1150	0.1241	−0.0035	0.1188	0.1244
QRH	0.1219	−0.0010	0.1181	0.1241	0.1173	−0.0008	0.1149	0.1186
(S-)M-2F-RHyp	0.1121	−0.0008	0.1099	0.1124	0.1115	−0.0006	0.1095	0.1123
(S-)M-2F-QHyp	0.1170	−0.0021	0.1140	0.1158	0.1188	−0.0029	0.1147	0.1168
(S-)M-2F-QRHyp	0.1164	−0.0031	0.1124	0.1142	0.1214	−0.0032	0.1169	0.1195
Sept. 4, 2012:								
RB-F	0.0739	−0.0007	0.0723	0.0739				
RB	0.0764	−0.0004	0.0750	0.0770				
2F-B	0.0781	−0.0016	0.0758	0.0774				
(S-)2F-RB	0.0746	−0.0011	0.0726	0.0742	0.0842	−0.0012	0.0813	0.0841
(S-)M-1F-RB	0.1172	−0.0400	0.0762	0.0805	0.0846	−0.0082	0.0755	0.0773
(S-)M-2F-RB	0.0965	−0.0091	0.0821	0.1062	0.1385	−0.0435	0.0833	0.1506
QRH	0.0775	−0.0006	0.0757	0.0785	0.0741	−0.0007	0.0724	0.0749
(S-)M-2F-RHyp	0.0739	−0.0008	0.0722	0.0738	0.0713	−0.0009	0.0697	0.0711
(S-)M-2F-QHyp	0.0815	−0.0031	0.0776	0.0792	0.0773	−0.0022	0.0746	0.0758
(S-)M-2F-QRHyp	0.0765	−0.0013	0.0746	0.0758	0.0766	−0.0011	0.0748	0.0763
Jan. 13, 2017:								
RB-F	0.0398	−0.0006	0.0386	0.0398				
RB	0.0386	−0.0004	0.0378	0.0387				
2F-B	0.0417	−0.0010	0.0399	0.0411				
(S-)2F-RB	0.0423	−0.0013	0.0404	0.0417	0.0456	−0.0015	0.0430	0.0459
(S-)M-1F-RB	0.1703	−0.1246	0.0403	0.0863	0.0954	−0.0542	0.0390	0.0627
(S-)M-2F-RB	0.2083	−0.1664	0.0386	0.1199	0.1693	−0.1195	0.0487	0.0734
QRH	0.0470	−0.0035	0.0414	0.0594	0.0436	−0.0016	0.0409	0.0446
(S-)M-2F-RHyp	0.0376	−0.0006	0.0365	0.0377	0.0390	−0.0005	0.0374	0.0397
(S-)M-2F-QHyp	0.0412	−0.0013	0.0396	0.0403	0.0408	−0.0010	0.0394	0.0401
(S-)M-2F-QRHyp	0.0411	−0.0013	0.0394	0.0402	0.0417	−0.0015	0.0399	0.0405

Notes: In the columns with numbers, we show first values of K_{VS} computed with deterministic methods (denoted \hat{K}_{VS}^{det}), then differences between the medians of \hat{K}_{VS}^{mc} and the values of \hat{K}_{VS}^{det} , lastly the percentiles 5 and 95 of \hat{K}_{VS}^{mc} . Statistics for \hat{K}_{VS}^{mc} are based on 100 samples. The cases where we have deemed estimation to difficult are highlighted in grey.

For the log-normal mixture models, we report only a single error in table 6 and that for M-2F-RB calibrated on October 14, 2011. As it nests the two-factor Bergomi model, the error is unsurprisingly low at 3 bps. For the hyperbolic models, we note that M-2F-QHyp and M-2F-QRHyp performs well on all three dates with errors of 4, 12, and 20 bps, respectively, 9, 5, and 22 bps. The other hyperbolic model M-2F-RHyp calibrates worst among the two-factor models; its errors are 27, 15, and 65 bps. Given that M-2F-QRHyp calibrates much better than M-2F-RHyp but uses the same kernel parameterisations, the poorer results for the latter model likely reflects a distributional problem and not a lack of flexibility for the volatility autocorrelations.

Our results show that substantial gains can be made with an extra factor, especially when volatility is low. If we compare 2F-B and M-2F-QHyp in figure 12, we see that they both achieve essentially perfect fits on October 14, 2011, and September 4, 2012. We can report that 2F-RB and M-2F-QRHyp likewise do although we, as noted, leave out the fits for visibility. However, on January 13, 2017, differences emerge. Even if 2F-B reproduces some of the decoupled term

structures of skew and curvature on this date, there is yet a very visible residual error; see figure 13 in particular. In contrast, M-2F-QHyp attains an almost perfect fit on this date also. While not shown, we can report that also on January 13, 2017, are the volatility smiles for M-2F-QRHyp, respectively, 2F-RB, almost indistinguishable from those of M-2F-QHyp, respectively, 2F-B.

Since the model M-2F-QHyp, respectively, M-2F-QRHyp, uses the same kernel parameterisations as the model 2F-B, respectively, 2F-RB, but calibrates almost perfectly on the low volatility date, we find it unlikely that the poorer fits for 2F-B and 2F-RB result from a lack of flexibility for the volatility autocorrelations. Rather we believe the problem is distributional and possibly is related to the observation that the log-normal distribution is too light-tailed; c.f. the commentary on page 17. Another possible explanation is that the poorer fits result from a misspecified elasticity of variance-of-variance which part one of our analysis suggested for the one-factor rough Bergomi model. We can expect this to be more problematic when the term structure of volatility is more pronounced as it then presumably will be more difficult to

Table 9. Calibration to SPX and VIX options.

S-2F-RB	θ	η	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	c	$\bar{\xi}_0$							Error	
2011-10-14	0.93	2.50	−0.70	−0.94	0.41	106.66	0.83	0.00	0.00	0.0197	0.0940							84	
2012-09-04	0.86	2.57	−0.59	−0.94	0.83	59.92	3.65	0.00	0.00	0.0134	0.0305							70	
2017-01-13	0.93	3.03	−0.39	−0.97	0.15	56.55	1.94	0.00	−0.19	0.0083	0.0157							118	
S-M-1F-RB	μ	η_1	η_2	ρ	λ	α	c	$\bar{\xi}_0$											Error
2011-10-14	0.65	0.73	2.59	−1.00	0.03	−0.30	0.0000	0.1038											197
2012-09-04	0.90	2.13	9.23	−0.97	1.41	−0.43	0.0134	0.0322											*
2017-01-13	0.69	1.99	7.68	−0.96	0.46	−0.43	0.0129	0.0469											*
S-M-2F-RB	μ	θ_1	θ_2	η_1	η_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	c	$\bar{\xi}_0$					Error
2011-10-14	0.33	0.79	0.93	0.79	2.80	−0.89	−0.99	0.83	141.87	0.90	0.00	0.00	0.0120	0.0957					88
2012-09-04	0.44	0.75	0.97	2.00	5.84	−0.51	−0.99	0.59	68.56	1.97	−0.05	−0.34	0.0150	0.0439					*
2017-01-13	0.24	0.28	0.34	1.30	8.48	−1.00	−0.85	0.86	174.34	2.88	−0.35	−0.05	0.0102	0.0777					*
QRH	a	b	c	ρ	λ	α	$\bar{\xi}_0$											Error	
2011-10-14	0.09	0.93	0.0075	1.00	0.16	−0.43	0.1550											54	
2012-09-04	0.07	0.74	0.0032	0.82	0.33	−0.44	0.3541											81	
2017-01-13	0.07	0.23	0.0009	0.94	2.21	−0.46	0.0417											89	
S-M-2F-RHyp	μ	θ_1	θ_2	η_1	η_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	c	$\bar{\xi}_{1,0}$	$\bar{\xi}_{2,0}$				Error
2011-10-14	0.13	1.00	0.10	5.10	7.14	−0.98	−0.56	0.36	0.00	1.05	−0.43	0.00	0.0000	−3.58	−0.99				157
2012-09-04	0.26	0.77	0.15	18.00	7.35	−1.00	−1.00	1.00	2.28	0.03	−0.43	−0.04	0.0023	−4.85	−3.69				78
2017-01-13	0.93	0.00	0.71	48.36	19.98	−0.98	−0.58	0.42	2.64	1.42	−0.40	−0.01	0.0076	−3.06	−2.35				120
S-M-2F-QHyp	μ	θ_1	θ_2	η_1	η_2	d_1	d_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	c	$\bar{\xi}_{1,0}$	$\bar{\xi}_{2,0}$				Error
2011-10-14	0.31	0.06	0.11	0.54	0.98	3.47	14.80	0.96	0.70	0.46	1.46	118.55	0.0112	−0.71	−0.10				73
2012-09-04	0.56	0.17	0.06	0.72	0.84	5.37	7.19	1.00	0.48	0.48	4.78	159.85	0.0081	−0.42	0.01				75
2017-01-13	0.02	0.31	0.04	1.41	0.73	22.47	6.24	0.88	0.72	0.32	3.66	98.43	0.0000	−0.73	0.01				40
S-M-2F-QRHyp	μ	θ_1	θ_2	η_1	η_2	d_1	d_2	ρ_{12}	ρ_{13}	ρ_{23}	λ_1	λ_2	α_1	α_2	c	$\bar{\xi}_{1,0}$	$\bar{\xi}_{2,0}$	Error	
2011-10-14	0.95	0.10	0.07	1.11	0.71	20.96	3.38	0.96	0.73	0.53	1.29	115.49	0.00	−0.02	0.0137	−0.58	−0.14	79	
2012-09-04	0.19	0.29	0.23	1.04	0.74	9.90	6.53	1.00	0.46	0.45	1.03	22.87	−0.16	−0.25	0.0095	−1.37	0.01	61	
2017-01-13	0.95	0.04	0.43	1.01	0.86	11.05	5.64	0.80	0.68	0.10	3.46	104.31	−0.01	−0.04	0.0009	−0.49	0.00	46	

Notes: Bar-notation denotes average values till 1 month. Errors are reported in basis points, though, some are left out where price estimation is deemed too difficult; we refer to the main text.

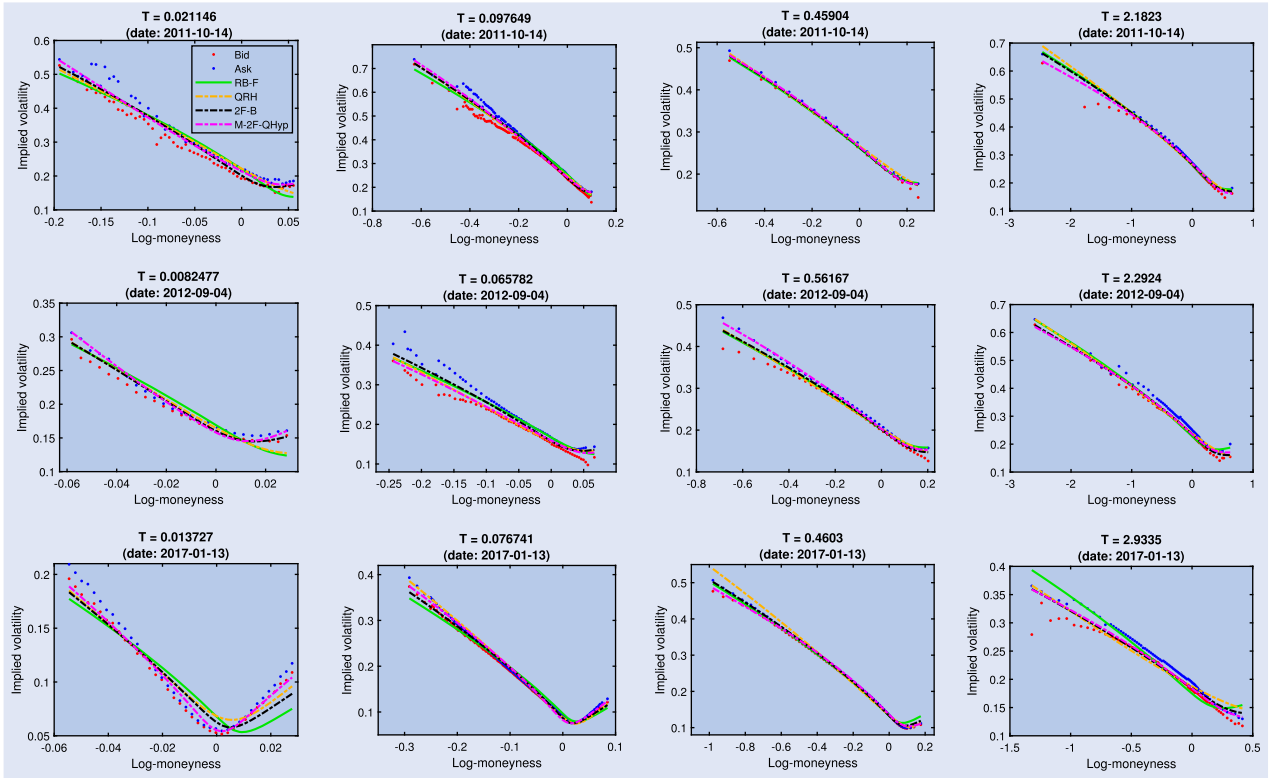


Figure 12. Calibration to SPX options (selected models and expiries).

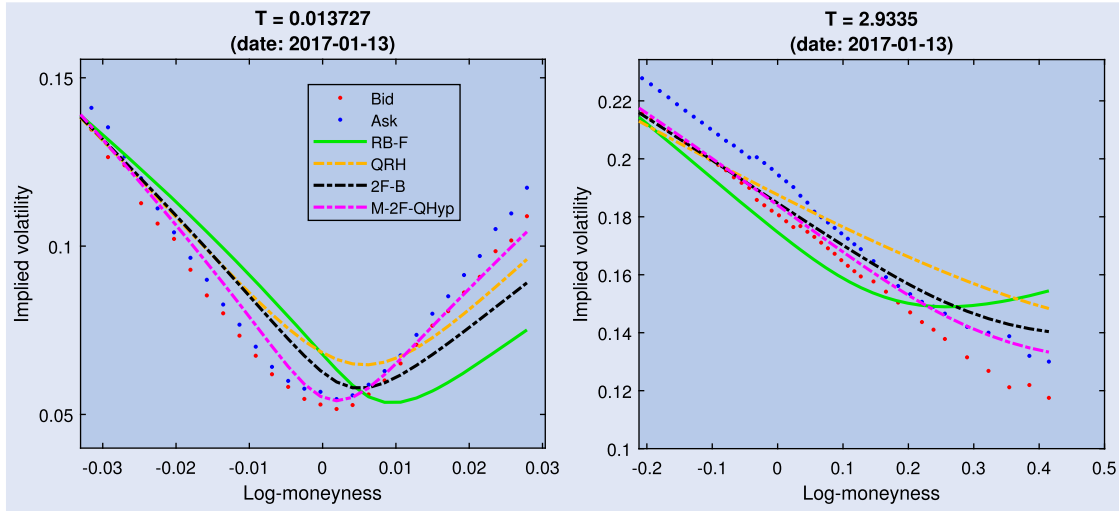


Figure 13. Calibration to SPX options (zoomed examples).

get the effective volatility-of-volatility right at both short and long time horizons. Indeed, we find that the volatility term structure is more marked on January 13, 2017: the values of $\frac{1}{T} \int_0^T \xi_0(t) dt$ with $T = 1$ (1 year) divided by those with $T = \frac{5}{252}$ (1 week) are 2.14, 2.49, and 5.82 on respectively October 14, 2011, September 4, 2012, and January 13, 2017; the reported numbers are based on ξ_0 as calibrated for the model 2F-B.

It is worth highlighting again that 2F-RB and M-2F-QRHyp, which both use the gamma kernel, calibrate about equally well to each of their non-rough counterparts 2F-B and M-2F-QHyp. Since M-2F-QHyp calibrates almost perfectly on all three dates, the conclusion then appears to be that while we do need two volatility factors to calibrate to

SPX options, their autocorrelation structure need not be very complicated as even exponential kernels seemingly suffice. We believe this is a non-trivial observation given all the evidence for roughness, mostly notably in Gatheral *et al.* (2018), and also the generally optimistic view of rough volatility that one is inclined to take on based on existing literature. The apparent success and popularity of one-factor rough volatility is likely a consequence of the fact that the rough fractional kernel generates mean-reversion at multiple time-scales even if driven by a single Brownian motion and which is in contrast to classical one-factor models whose autocorrelation structure likely is much too simple. It is then somewhat meaningful that the differences disappear, or at least significantly diminish, when we consider two-factor volatility models where

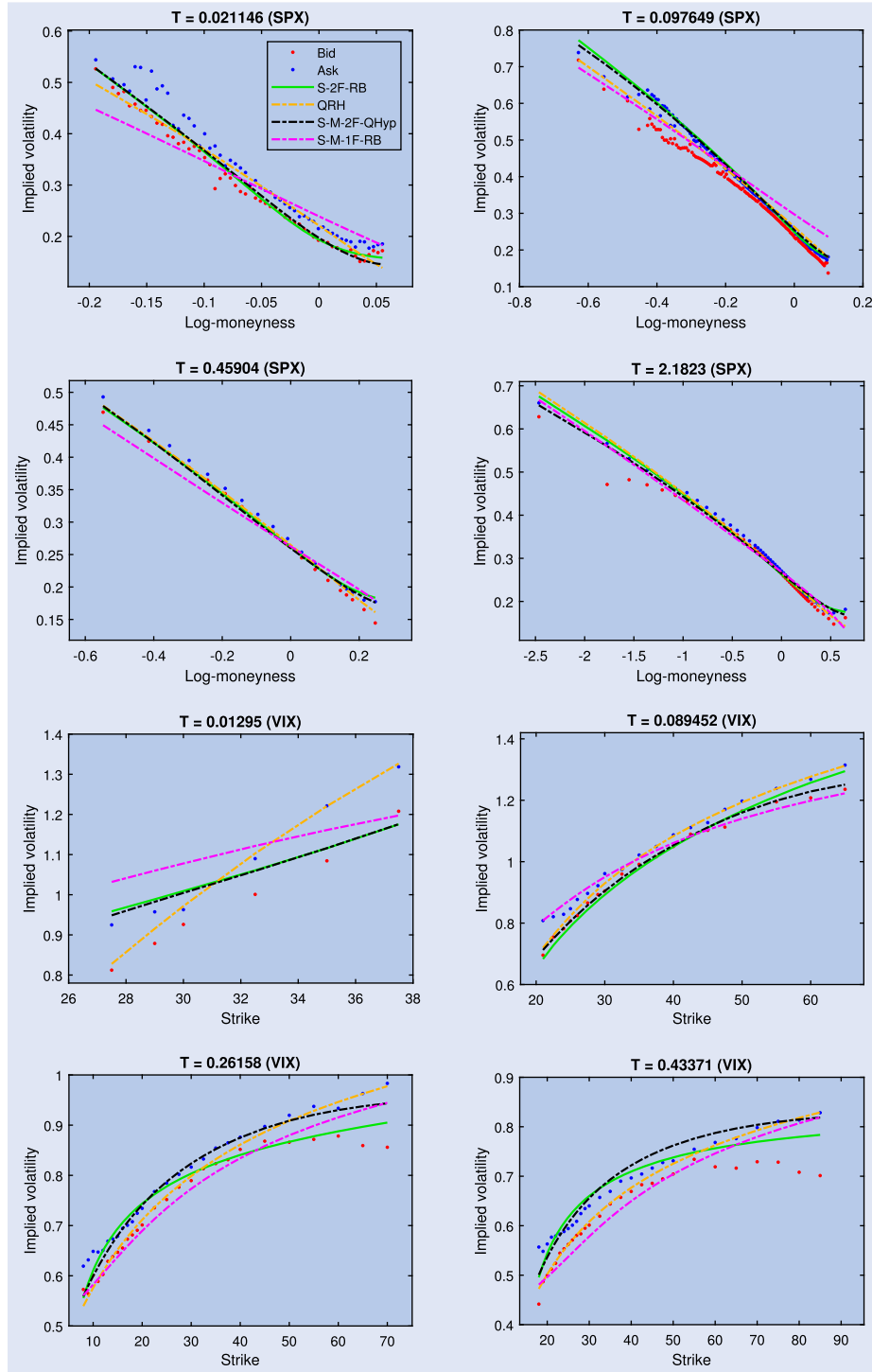


Figure 14. Joint calibration to SPX and VIX options (selected models and expiries). Date: 2011-10-14.

even non-rough ones allow mean-reversion at both short and long time-scales. However, as two factors seemingly are needed in any case to capture the decoupled nature of the volatility surface, and because computational costs are much worse under rough volatility, we find it hard to justify modelling volatility rough based on our results so far. As the reader will see, the remaining results are in line with this conclusion.

Before we move on to the joint calibrations, let us though stress that we cannot from our results conclude whether or not volatility *truly* is rough as that can only be finally answered

in the asymptotic limit of very small time-scales. Rather the point to be made is that, when viewed in a two-factor framework, roughness does not appear to matter at the time-scales relevant to the pricing of SPX options and, as the reader will see, VIX options. We note that the debate of roughness in many ways resembles that of long-memory which deals with the other asymptotic limit and for which the authors of Gatheral *et al.* (2018) make a comparable observation noting that while their Rough Fractional Stochastic Volatility (RFSV) model does not truly display long-memory as is found in many works (see e.g. Bennedsen *et al.* 2021) common statistical

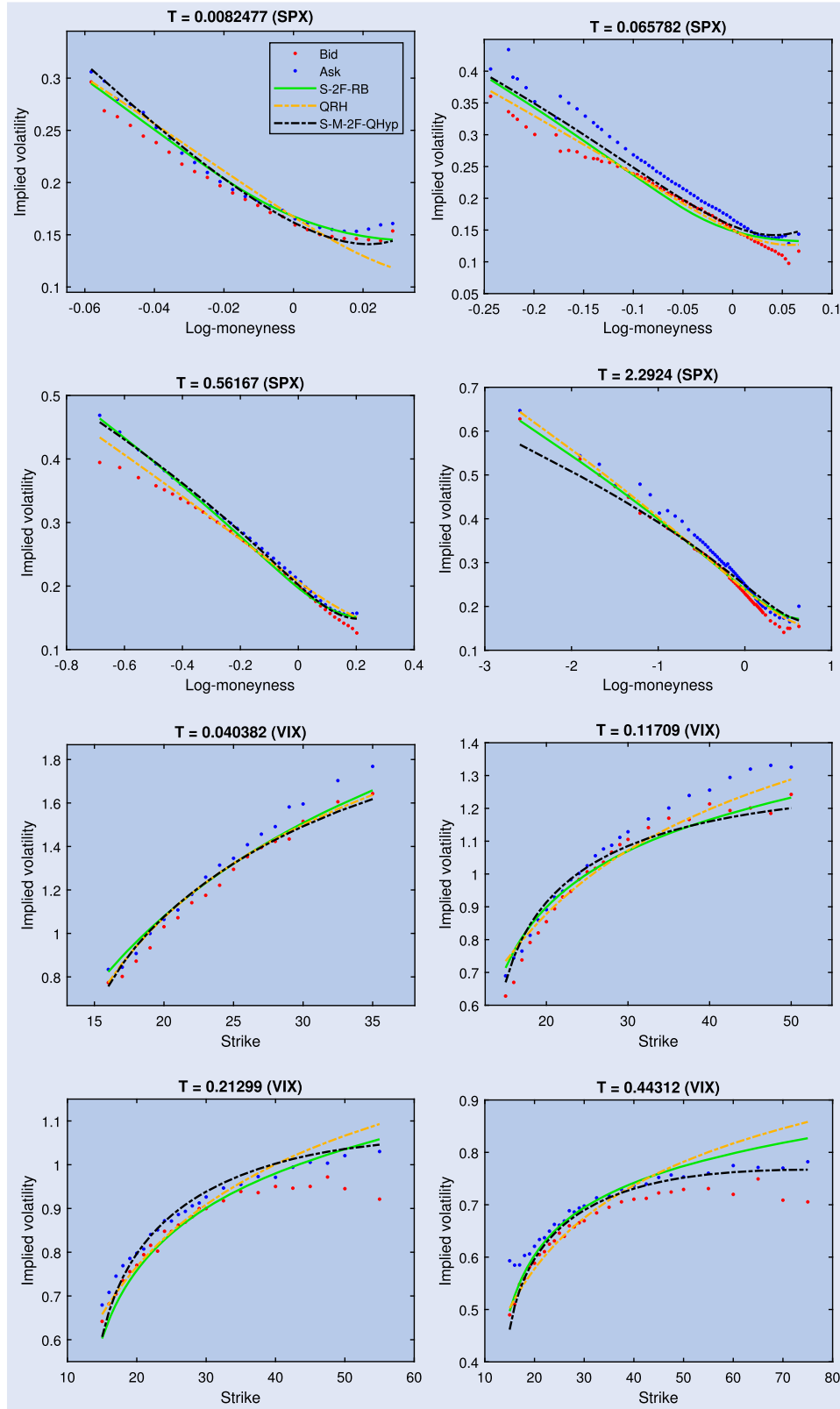


Figure 15. Joint calibration to SPX and VIX options (selected models and expiries). Date: 2012-09-04.

tests applied to sample paths under the model may indeed result in false positive detections of it.

In what follows, we discuss the joint calibrations. We focus on the models with log-normal terms first and start with October 14, 2011, where errors are reported for them all. With S-M-1F-RB we get an error of 197 bps which is significantly

worst. The reason is evident from figure 14, where we note that although the VIX smiles are fitted reasonably, too little skew is produced for short term SPX options. We take it as evidence that the model does not sufficiently decouple the variability of the spot variance (which drives S_t) from the remainder of the forward variance curve on which the VIX

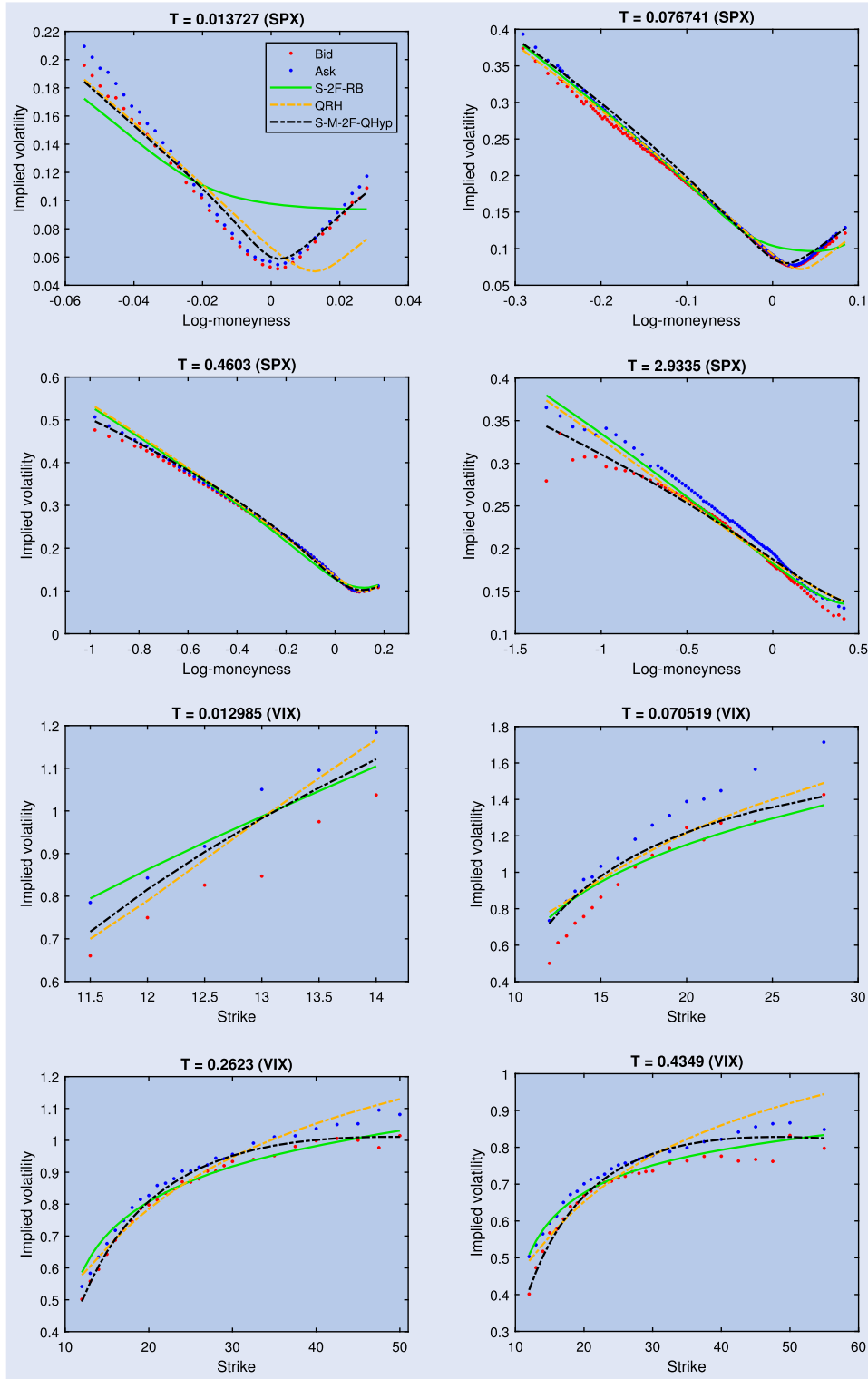


Figure 16. Joint calibration to SPX and VIX options (selected models and expiries). Date: 2017-01-13.

index depends. We note that the model struggles to reconcile both markets despite that it calibrates a correlation parameter of $\rho = -1^\dagger$ and uses the gamma kernel which nests the rough fractional kernel and allows mean-reversion at multiple time-scales. We find it unlikely that the problem arises from the inability of the model to create a term structure of

$^\dagger \rho$ allows us to influence the SPX skews independent of the VIX smiles.

(a)symmetry of which there is little on this date, or that the problem is distributional as the VIX smiles are fitted fairly well. In line with the results of section 4.3, a more plausible explanation is indeed that the gamma kernel, the fractional kernel especially, as used in this model, lacks flexibility in decoupling the short and long lag volatility autocorrelations. ‡

‡ Note that the volatility autocorrelation (or mean-reversion) structure tends to influence the relative variability of $V_t = \xi_t(t)$ vs. $\xi_t(u)$ for larger $u - t$. The rough Bergomi dynamics $d\xi_t(u)/\xi_t(u) =$

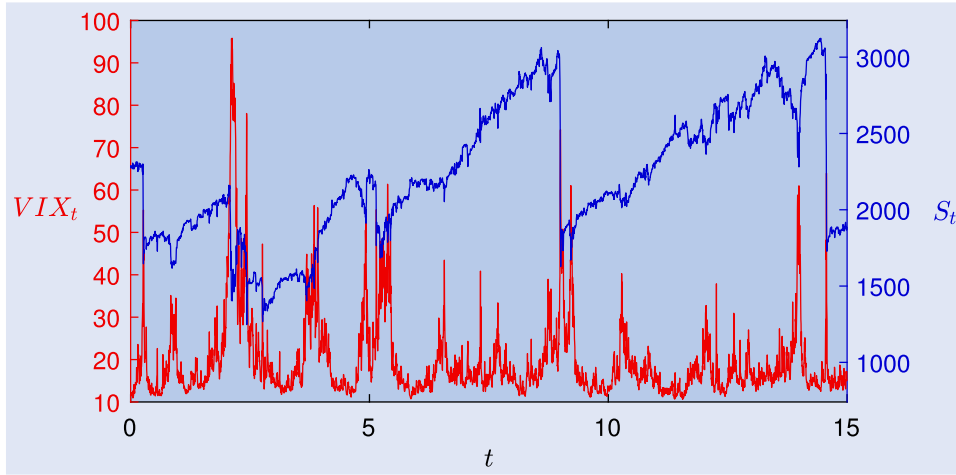


Figure 17. Simulation of S-M-2F-QHyp as calibrated to SPX and VIX options on January 13, 2017. The curves $\zeta_{1,0}$ and $\zeta_{2,0}$ are extended flat and S_t is simulated without drift.

The same problem is reflected in Guyon (2018) where a similar ‘skewed’ rough Bergomi model is calibrated.

The models S-2F-RB and S-M-2F-RB perform much better with roughly half the errors at 84 and 88 bps. In figure 14, we see that the fit for S-2F-RB also visually is very decent. We do not show S-M-2F-RB in the plot but can report that the fit is comparable to that of S-2F-RB. For September 4, 2012, we only report errors for S-2F-RB among the log-normal type models. It though again performs well with an error of 70 bps and a very decent fit (figure 15). It is worth noting that S-2F-RB calibrated to both of these dates, and S-M-2F-RB calibrated to October 14, 2011, all have $\alpha_1 = \alpha_2 = 0$ so they in fact are two-factor classical models. This is more evidence that only two exponential terms are needed to capture the relevant volatility autocorrelations.

On January 13, 2017, we, among the log-normal type models, again only report an error for S-2F-RB. However, on this date, it calibrates much worse with an error of 118 bps and a notably poor fit (figure 16). While the VIX smiles are fitted fairly well, the short term SPX smiles are too flat. We believe the problem is distributional because S-M-2F-QRHyp, which is driven by the same two Gaussian processes, attains an error of only 46 bps; we discuss the hyperbolic models more momentarily. The problem is reflected in the shift parameter that is calibrated to $c = 0.0083$ (9.11% in volatility terms) and which is high relative to the volatility level (the VIX index is at 11.23). This indicates that S-2F-RB relies too much on the volatility shift to skew the VIX smiles. As shown in figure 11, that can precisely flatten the SPX skews. Since S-2F-RB calibrates poorly on the low volatility date and the log-normal mixture type models have estimation problems, we conclude that the log-normal distribution, all in all, is ill-suited for the joint calibration problem.

Next, we consider the quadratic rough Heston (QRH) model which, with errors of 54, 81, and 89 bps, performs

better overall compared to S-2F-RB and S-M-2F-RB. Figures 14–16 show that it captures both the SPX *skews* and the VIX *smiles*; that is, besides the very short term SPX skew on January 13, 2017. This is impressive given that the model only has six parameters besides the curve ζ_0 and only one Brownian motion driving volatility. It is somewhat surprising that it performs so much better than S-M-1F-RB on October 14, 2011, (the only date we can compare on) which is also one-factor with a gamma kernel. We suspect that QRH calibrates fairly well to the joint problem because of the squaring operation. More precisely, we believe it might increase mean-reversion by amplifying the relative differences between small and large values thereby helping to decouple the SPX skews from the level of the VIX implied volatilities. However, regardless of this, the model is yet unable to reproduce the term structure of smile (a)symmetry that is very visible on January 13, 2017. We conclude that the model, perhaps, is a bit *too* simple.

Lastly, we consider the hyperbolic models. We note first that S-M-2F-RHyp calibrates poorly overall with errors of 157, 78, and 120 bps. This is unsurprising given the also somewhat unsatisfactory calibrations of M-2F-RHyp to SPX options. Much better results are instead achieved with the quadratic but non-rough model S-M-2F-QHyp which results in the errors 73, 75, and 40 bps. Figures 14–16 also show that the fits for S-M-2F-QHyp are very decent visually on all three dates. Compared to quadratic rough Heston, we note that the fit is just slightly worse on the high volatility date, though still fairly decent, but notably better on the low volatility date. The rough version S-M-2F-QRHyp performs similarly with the errors 79, 61, and 46 bps.

The very decent and robust calibrations of S-M-2F-QHyp suggest that the joint calibration problem is largely solvable with classical two-factor volatility. This goes against the idea that jumps or roughness are needed to calibrate jointly. Other works in the literature contain similar conclusions in that simpler models may do equally well to, or even better than, more sophisticated and complex ones. We especially note this in relation to jump models. In Kaeck and Alexander (2013) the authors e.g. find that a simple GARCH diffusion model outperforms jump diffusions.

$\eta\sqrt{2H}(u-t)^{H-\frac{1}{2}}dW_{2,t}$ illustrates this point; note the influence of H on the volatility term structure of $\xi_t(\cdot)$. The poor joint calibration of S-M-1F-RB suggest that even the gamma kernel is not flexible enough to model the term structure of volatility of $\xi_t(\cdot)$, at least for this type of model.

As the analysis of the present section is based on a somewhat small sample size, the conclusion should nevertheless be taken with some moderation. We leave a more extensive historical calibration of S-M-2F-QHyp and the other models for future work.

We end our discussion with a simulation of S-M-2F-QHyp as calibrated to the joint problem on January 13, 2017; we refer to figure 17 where we show a trajectory of S_t and VIX_t across 15 years. Despite that volatility is driven by two Ornstein-Uhlenbeck processes, the model is able to generate jump-like behaviour for the S&P 500 index characteristic of rough volatility. The trajectory of the VIX index also qualitatively resembles the historical trajectory of figure 4.

6. Conclusion

We have in this paper shown that the one-factor rough Bergomi and rough Heston models fail to create a pronounced enough term structure of smile effect for SPX options. Moreover, we have discovered that short expiry SPX smiles generally are more symmetric than long expiry smiles, a feature we neither found that these models could recreate. To calibrate to SPX options with a continuous-path stochastic volatility model our results suggests that two factors are needed for volatility, three for the S&P 500 index. One volatility factor should be less negatively correlated with the index, more noisy, and mean-revert faster. Once extended to a two-factor volatility framework, we found no meaningful differences between comparable rough and classical non-rough models and obtained decent fits with even a two-factor Bergomi model. The two-factor Bergomi model was though still not fully aligned with SPX prices on a date with low volatility. Moreover, we were for several reasons unable to calibrate, across volatility levels, jointly to SPX and VIX options with more general models based on the log-normal distribution. We therefore considered several alternatives, including a two-factor volatility model based on the quadratic and hyperbolic transformations and which is driven by two ordinary Ornstein-Uhlenbeck processes. We demonstrated that this model can calibrate almost perfectly to SPX options and very decently to the joint problem, also across volatility levels. From this we conclude that the SPX and VIX options markets can largely be reconciled with classical two-factor volatility, all without roughness and jumps.

Acknowledgments

The author is thankful for the valuable comments provided by the anonymous referees which led to a much improved paper.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Sigurd Emil Rømer  <http://orcid.org/0000-0001-8990-0483>

References

- Abi Jaber, E., Lifting the Heston model. *Quant. Finance*, 2019, **19**(12), 1995–2013.
- Abi Jaber, E. and El Euch, O., Markovian structure of the Volterra Heston model. *Stat. Probab. Lett.*, 2019a, **149**, 63–72.
- Abi Jaber, E. and El Euch, O., Multifactor approximation of rough volatility models. *SIAM J. Financ. Math.*, 2019b, **10**(2), 309–349.
- Abi Jaber, E., Larsson, M. and Pulido, S., Affine Volterra processes. *Ann. Appl. Probab.*, 2019, **29**(5), 3155–3200.
- Alòs, E., García-Lorite, D. and Muguruza, A., On smile properties of volatility derivatives: Understanding the VIX skew. *SIAM J. Financ. Math.*, 2022, **13**(1), 32–69.
- Bayer, C. and Stemper, B., Deep calibration of rough stochastic volatility models. Working Paper, 2018, arXiv:1810.03399 (accessed 27 July 2021).
- Bayer, C., Friz, P. and Gatheral, J., Pricing under rough volatility. *Quant. Finance*, 2016, **16**(6), 887–904.
- Bayer, C., Friz, P.K., Gulisashvili, A., Horvath, B. and Stemper, B., Short-time near-the-money skew in rough fractional volatility models. *Quant. Finance*, 2019, **19**(5), 779–798.
- Bayer, C., Friz, P.K., Gassiat, P., Martin, J. and Stemper, B., A regularity structure for rough volatility. *Math. Financ.*, 2020, **30**(3), 782–832.
- Bennedsen, M., Lunde, A. and Pakkanen, M.S., Decoupling the short- and long-term behavior of stochastic volatility. *J. Financ. Econ.*, 2021, **nbaa049**, 1–46. <https://doi.org/10.1093/jfinc/ebaa049>.
- Bennedsen, M., Lunde, A. and Pakkanen, M.S., Decoupling the short- and long-term behavior of stochastic volatility. Working Paper, 2017a, arXiv: 1610.00332v2 (accessed 27 July 2021).
- Bennedsen, M., Lunde, A. and Pakkanen, M.S., Hybrid scheme for Brownian semistationary processes. *Finance Stoch.*, 2017b, **21**, 931–965.
- Bergomi, L., Smile dynamics III. *Risk*, October 2008, 90–96.
- Bergomi, L. and Guyon, J., Stochastic volatility's orderly smiles. *Risk*, May 2012, 60–66.
- Bishop, C.M., *Pattern Recognition and Machine Learning*, 2006 (Springer: New York).
- Callegaro, G., Grasselli, M. and Pagès, G., Fast hybrid schemes for fractional Riccati equations (rough is not so tough). *Math. Oper. Res.*, 2021, **46**(1), 221–254.
- Chung, S.-L., Tsai, W.-C., Wang, Y.-H. and Weng, P.-S., The information content of the S&P 500 index and VIX options on the dynamics of the S&P 500 index. *J. Futures Mark.*, 2011, **31**(12), 1170–1201.
- Cont, R. and Kokholm, T., A consistent pricing model for index options and volatility derivatives. *Math. Financ.*, 2013, **23**(2), 248–274.
- De Marco, S., Volatility derivatives in (rough) forward variance models. Stochastic Analysis and its Applications, Oaxaca, May 2018.
- Diethelm, K., Ford, N.J. and Freed, A.D., Detailed error analysis for a fractional Adams method. *Numer. Algorithms*, 2004, **36**(1), 31–52.
- El Euch, O. and Rosenbaum, M., Perfect hedging in rough Heston models. *Ann. Appl. Probab.*, 2018, **28**(6), 3813–3856.
- El Euch, O., Fukasawa, M. and Rosenbaum, M., The microstructural foundations of leverage effect and rough volatility. *Finance Stoch.*, 2018, **22**(2), 241–280.
- El Euch, O., Fukasawa, M., Gatheral, J. and Rosenbaum, M., Short-term at-the-money asymptotics under stochastic volatility models. *SIAM J. Financ. Math.*, 2019a, **10**(2), 491–511.

- El Euch, O., Gatheral, J. and Rosenbaum, M., Roughening Heston. *Risk*, May 2019b, 84–89.
- Forde, M. and Smith, B., Rough Heston with jumps - Joint calibration to SPX/VIX level and skew as $T \rightarrow 0$, and issues with the quadratic rough Heston model, 2021. Available online at: <https://nms.kcl.ac.uk/martin.forde/> (accessed 25 April 2022).
- Forde, M. and Zhang, H., Asymptotics for rough stochastic volatility models. *SIAM J. Financ. Math.*, 2017, **8**(1), 114–145.
- Forde, M., Gerhold, S. and Smith, B., Small-time, large-time, and $H \rightarrow 0$ asymptotics for the rough Heston model. *Math. Financ.*, 2021a, **31**(1), 203–241.
- Forde, M., Smith, B. and Viitasari, L., Rough volatility and CGMY jumps with a finite history and the rough Heston model – Small-time asymptotics in the $k\sqrt{t}$ regime. *Quant. Finance*, 2021b, **21**(4), 541–563.
- Forde, M., Fukasawa, M., Gerhold, S. and Smith, B., The Riemann-Liouville field and its GMC as $H \rightarrow 0$, and skew flattening for the rough Bergomi model. *Stat. Probab. Lett.*, 2022, **181**, 109265.
- Friz, P.K., Gassiat, P. and Pigato, P., Precise asymptotics: Robust stochastic volatility models. *Ann. Appl. Probab.*, 2021, **31**(2), 896–940.
- Friz, P.K., Gassiat, P. and Pigato, P., Short-dated smile under rough volatility: Asymptotics and numerics. *Quant. Finance*, 2022, **22**(3), 463–480.
- Fukasawa, M., Short-time at-the-money skew and rough fractional volatility. *Quant. Finance*, 2017, **17**(2), 189–198.
- Gatheral, J., *The Volatility Surface: A Practitioner's Guide*, 2006 (John Wiley & Sons: Hoboken, NJ).
- Gatheral, J., Consistent modeling of SPX and VIX options. The Fifth World Congress of the Bachelier Finance Society, London, 18 July, 2008.
- Gatheral, J. and Jacquier, A., Arbitrage-free SVI volatility surfaces. *Quant. Finance*, 2014, **14**(1), 59–71.
- Gatheral, J. and Radoičić, R., Rational approximation of the rough Heston solution. *Int. J. Theor. Appl. Finance*, 2019, **22**(3), 1950010.
- Gatheral, J., Jaisson, T. and Rosenbaum, M., Volatility is rough. *Quant. Finance*, 2018, **18**(6), 933–949.
- Gatheral, J., Jusselin, P. and Rosenbaum, M., The quadratic rough Heston model and the joint S&P 500/Vix smile calibration problem. *Risk*, May 2020.
- Goodfellow, I., Bengio, Y. and Courville, A., *Deep Learning*, 2016 (MIT Press). Available online at: <http://www.deeplearningbook.org> (accessed 25 April 2022).
- Gulisashvili, A., Time-inhomogeneous Gaussian stochastic volatility models: Large deviations and super roughness. *Stoch. Process. Appl.*, 2021, **139**, 37–79.
- Guo, I., Loeper, G., Oblój, J. and Wang, S., Joint modelling and calibration of SPX and VIX by optimal transport. *SIAM J. Financ. Math.*, 2022, **13**(1), 1–31.
- Guyon, J., On the joint calibration of SPX and VIX options. *Quant-Minds*, Lisbon, 16 May, 2018.
- Guyon, J., Inversion of convex ordering in the VIX market. *Quant. Finance*, 2020, **20**(10), 1597–1623.
- Guyon, J., The VIX future in Bergomi models: Fast approximation formulas and joint calibration with S&P 500 skew. Working Paper, 2022, SSRN:3720315 (accessed 8 May 2022).
- Hagan, P.S., Kumar, D., Lesniewski, A.S. and Woodward, D.E., Managing smile risk. *Wilmott Mag.*, September 2002, 84–108.
- Heston, S.L., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 1993, **6**(2), 327–343.
- Hornik, K., Approximation capabilities of multilayer feedforward networks. *Neural Netw.*, 1991, **4**(2), 251–257.
- Horvath, B., Jacquier, A. and Tankov, P., Volatility options in rough volatility models. *SIAM J. Financ. Math.*, 2020, **11**(2), 437–469.
- Horvath, B., Muguruza, A. and Tomas, M., Deep learning volatility: A deep neural network perspective on pricing and calibration in (rough) volatility models. *Quant. Finance*, 2021, **21**(1), 11–27.
- Jacquier, A., Martini, C. and Muguruza, A., On VIX futures in the rough Bergomi model. *Quant. Finance*, 2018, **18**(1), 45–61.
- Kaeck, A. and Alexander, C., Stochastic volatility jump-diffusions for European equity index dynamics. *Eur. Financ. Manag.*, 2013, **19**(3), 470–496.
- Kahl, C. and Jäckel, P., Fast strong approximation Monte Carlo schemes for stochastic volatility models. *Quant. Finance*, 2006, **6**(6), 513–536.
- Kidger, P. and Lyons, T., Universal approximation with deep narrow networks. *Proc. Mach. Learn. Res.*, 2020, **125**, 2306–2327.
- Kingma, D.P. and Ba, J., Adam: A method for stochastic optimization. Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015, arXiv:1412.6980 (accessed 27 July 2021).
- Kokholm, T. and Stisen, M., Joint pricing of VIX and SPX options with stochastic volatility and jump models. *J. Risk Finance*, 2015, **16**(1), 27–48.
- Lacombe, C., Muguruza, A. and Stone, H., Asymptotics for volatility derivatives in multi-factor rough volatility models. *Math. Finan. Econ.*, 2021, **15**, 545–577.
- Lewis, A.L., A simple option formula for general jump-diffusion and other exponential Levy processes. Working Paper, 2001, SSRN:282110 (accessed 27 July 2021).
- Lord, R. and Kahl, C., Optimal Fourier inversion in semi-analytical option pricing. Tinbergen Institute Discussion Paper No. 2006-066/2, 2007, SSRN:921336 (accessed 27 July 2021).
- McCrickerd, R. and Pakkanen, M.S., Turbocharging Monte Carlo pricing for the rough Bergomi model. *Quant. Finance*, 2018, **18**(11), 1877–1886.
- McGhee, W.A., An artificial neural network representation of the SABR stochastic volatility model. Working Paper, 2018, SSRN:3288882 (accessed 27 July 2021).
- Pacati, C., Renò, R. and Santilli, M., Heston model: Shifting on the volatility surface. *Risk*, November 2014, 54–59.
- Pacati, C., Pompa, G. and Renò, R., Smiling twice: The Heston++ model. *J. Bank. Financ.*, 2018, **96**, 185–206.
- Roper, M., Arbitrage free implied volatility surfaces. Preprint, School of Mathematics and Statistics, The University of Sydney, Australia, 2010.
- Rømer, S.E., Hybrid multifactor scheme for stochastic Volterra equations with completely monotone kernels. Working Paper, 2022, SSRN:3706253 (accessed 5 May 2022).
- Rømer, S.E. and Poulsen, R., How does the volatility of volatility depend on volatility? *Risks*, 2020, **8**(2), 59.
- Scott, L.O., Option pricing when the variance changes randomly: Theory, estimation, and an application. *J. Financ. Quant. Anal.*, 1987, **22**(4), 419–438.