



August 6th, 2018

Rough Volatility

Option Pricing Under the Rough Bergomi Model

Author: Sigurd Emil Rømer
Email: htq530@alumni.ku.dk

Master Thesis in Mathematics - Economics

Advisor: Rolf Poulsen
Email: rolf@math.ku.dk

Abstract

In this thesis we investigate volatility on the S&P 500 using high frequency trade data on the SPY. We show that log-volatility is a rough unifractal process that is approximately Gaussian when aggregated. Based on these findings we propose to model volatility via a fractional Brownian motion with Hurst exponent H less than $\frac{1}{2}$. Under a deterministic change of measure this leads to the rough Bergomi pricing model. For this model we investigate methods for speeding up simulation and present a fast calibration scheme using various approximations. Calibrating to SPX option prices we find that the model is able to capture the power-law decay of skew observed in practise. We furthermore consider hedging in the model on forward variance curve form and relate the static properties of a power-law decay of skew to the dynamic properties as measured through the skew-stickiness-ratio (SSR). We find that the model with empirically relevant values of H is able to capture the typically observed SSR values of around 1.5. Finally we present some initial thoughts on how one can bet on H in the rough Bergomi model.

Keywords: Volatility Modelling; Rough Volatility; Fractional Brownian Motion; Implied Volatility Surface; Option Pricing; Rough Bergomi Model; Skew-Stickiness-Ratio

Contents

1	Introduction	1
2	Empirical Facts of Volatility	3
2.1	Estimating Volatility from High-Frequency Data	3
2.2	Empirical Properties of Volatility	9
3	Pricing Model	20
3.1	Model Under P	20
3.2	Volatility Process Under Q	22
3.3	Full Model Under Q	23
4	Simulation	26
4.1	A Simple Riemann Scheme	26
4.2	Exact Simulation	27
4.3	Hybrid Scheme	28
4.4	Variance Reduction	31
5	Calibration	35
5.1	Estimating The Forward Variance Curve	35
5.2	Market Data and Implied Yields	38
5.3	SVI Parameterization	39
5.4	Calibrating the rough Bergomi model	43
6	Hedging	48
6.1	Hedging Strategy	48
6.2	The Skew-Stickiness-Ratio	52
6.3	Trading the Skew-Stickiness-Ratio	56
7	Conclusion	62
A	Numerical Results	63
B	Theory	81
	Bibliography	101

Chapter 1

Introduction

Ever since the publication of the famous Black-Scholes formula in 1973 a large chunk of research in mathematical finance has revolved around derivatives pricing. A revolutionary idea in the Black-Scholes model is that a contingent claim can be perfectly hedged by a simple dynamic trading strategy in the underlying asset (typically a stock). The simplicity of the hedging strategy is however largely a product of assuming the volatility of the underlying asset constant. In reality that volatility is stochastic. Hedging with the Black-Scholes model therefore introduces a volatility risk that needs to be managed. Furthermore we observe non-flat implied volatility smiles in practise which conflicts with the model. To solve these problems several stochastic volatility (SV) models have been developed. Of the more famous ones one can mention the Heston model. However, while such conventional SV models allow for a framework where volatility risk can be hedged, and while they produce non-flat implied volatility smiles, they are unable to properly capture 1) the rapidly increasing implied volatility skew for short expiries and 2) the dynamic properties of implied volatility. Jump-diffusion models have been proposed as a solution to (1) but generally fail miserably at capturing (2). This highlights the need for a new generation of pricing models. New research proposes rough volatility models as the next step, claiming that this class of models can solve both (1) and (2). It is exactly rough volatility models that we focus on in this thesis.

Let us now elaborate on the empirical facts (1) and (2). For that purpose consider first an expiry $T > 0$, strike K vanilla option with time-to-expiry $\tau > 0$ and log-moneyness $k := \log(K/F_T)$ where F_T is the expiry T forward price of the underlying. Let then $\sigma_{BS}(k, \tau)$ be the Black-Scholes implied volatility for such an option. Define now the at-the-money ($k = 0$) skew as

$$\psi(\tau) := \frac{\partial}{\partial k} \sigma_{BS}(k, \tau) \Big|_{k=0}. \quad (1.1)$$

To state the first fact, we observe in practise the power-law relationship $|\psi(\tau)| \approx A\tau^{-\gamma}$ and that across all expiries. Here A is some constant and typically $\gamma \in (0, \frac{1}{2})$ such that skew explodes for short expiries. Conventional SV models are unable to capture this relationship and therefore cannot properly fit an implied volatility surface for both short and long term expiries at the same time.¹ Secondly, for the purpose of hedging options the dynamics of the implied volatility surface produced by a hedging model is of central importance. To see this let $P_{BS}(t, S, \sigma)$ be the Black-Scholes formula for some expiry $T > 0$ vanilla option with time-to-expiry $\tau := T - t$, current stock price S and a Black-Scholes volatility of σ . The main component in a hedge portfolio for a typical option is the amount to hold in the underlying stock. This amount is commonly referred to as the *delta* and assuming the European option has log-moneyness k this can be computed as

$$\frac{dP_{BS}(t, S_t, \sigma_{BS}(k, \tau))}{dS} = \frac{\partial P_{BS}}{\partial S} + \frac{\partial P_{BS}}{\partial \sigma} \cdot \frac{d\sigma_{BS}(k, \tau)}{dS}. \quad (1.2)$$

¹As an example, under Heston skew converges to a constant for short expiries and long term it behaves as τ^{-1} .

The main model dependent quantity in the above is the factor $\frac{d\sigma_{BS}(k,\tau)}{dS}$. Depending on how a hedging model relates moves in implied volatilities to moves in the stock price this quantity will look different and you will therefore get different values of the delta. In (Bergomi 2009) a new summary measure of the joint dynamics of the stock and implied volatility is proposed; namely the so-called skew-stickiness-ratio (SSR). We define the SSR for expiry $T > 0$ as

$$R_T := \frac{1}{\psi(T)} \frac{E(d \log S d\hat{\sigma}_{F_T})}{E((d \log S)^2)} \quad (1.3)$$

where $\hat{\sigma}_{F_T}$ is the at-the-money implied volatility for expiry $T > 0$ vanilla options.² In works such as (Bergomi 2009) and (Bergomi 2016) the SSR has been estimated for Euro Stoxx 50 and the S&P 500 to be around 1.5 for all expiries. Conventional SV models are unfortunately also unable to achieve this value of SSR across all expiries.

The class of rough volatility models is supposed to solve both of the above problems and as is indicated in the name, the source of the solution turns out to be the notion of *roughness*. Intuitively roughness measures how jagged the path of a given process is; the opposite of a rough path being a smooth one. Under some assumptions roughness can be measured via the so-called Hurst exponent that is a number $H \in (0, 1)$. A lower value of H results in a more rough process and so we call the process rough if $H < \frac{1}{2}$. This is the sense in which rough volatility is to be understood.

A number of papers have in the recent years been published on the topic of rough volatility. Notably is (Gatheral et al. 2017) where the rough fractional stochastic volatility (RFSV) model is proposed. Here log-volatility is modelled via a so-called fractional Brownian motion with a Hurst exponent of $H < \frac{1}{2}$ thereby producing a rough process. This model is found to correspond well to the empirical stylized facts of volatility. Later in (Bayer et al. 2016) the specific rough Bergomi model was proposed. This particular model has been especially well researched and is the model we focus on in this thesis. Other authors have developed fast and accurate simulation schemes for the rough Bergomi model, see (Bennedsen et al. 2017b), and estimators of option prices, see (McCrickerd & Pakkanen 2018). Various expansion formulas have also been derived. An example is the short expiry expansions derived in (Fukasawa 2015) for a general class of rough volatility models. While we will consider many of the aspects of rough volatility described in the above mentioned articles we will make a few new contributions. Firstly, as we generally find calibration of the rough Bergomi model to be difficult and slow we propose a fast calibration method. Secondly, while trading the SSR has been explored in (Bergomi 2009) we have not found any literature considering trading the SSR in a rough volatility model. As we find the SSR to be connected to the Hurst exponent H , we present our initial thoughts and results on how one can trade H via the SSR in the rough Bergomi model. The thesis is structured as follows: In the next chapter, chapter 2, we consider the volatility process under the historical probability measure by using high frequency data on the SPY. We show that log-volatility is rough, unifractal and has increments that are aggregationally Gaussian. With these facts we decide to model volatility via a fractional Brownian motion with Hurst exponent $H < \frac{1}{2}$. In chapter 3 we then consider a specific change of measure resulting in the so-called rough Bergomi pricing model which we will consider in the rest of the thesis. In chapter 4 we consider simulation under the model and present a hybrid scheme that is both fast and accurate. We also briefly consider a few variance reduction methods for estimating European option prices. With these things settled we move on to calibration in chapter 5 where we calibrate the model to SPX options data. Here we also propose the fast calibration method. In chapter 6 we consider hedging in the model. In relation to this we consider the dynamic properties of the model as summarized by the SSR. We conclude the chapter by presenting some thoughts on how one via the SSR can make a bet on the Hurst exponent H in a misspecified market.

²We use F_T as a subscript to indicate that the strike of $\hat{\sigma}_{F_T}$ is dynamic across time such that log-moneyness is always zero. That is, $\hat{\sigma}_{F_T}$ is the expiry $T > 0$ implied volatility corresponding to the strike K s.t. $k := \log\left(\frac{K}{F_T}\right) = 0$.

Chapter 2

Empirical Facts of Volatility

In this chapter we will investigate the empirical properties of the volatility process of the SPY. For this purpose we will use high frequency tick-by-tick data obtained from the Trades and Quotes (TAQ) database. For the analysis we will primarily consider the log volatility process which we find exhibits some stable properties. These properties will be the foundation of the volatility model we will formally state in the next chapter.

The chapter is structured as follows: First we explain how to estimate integrated variance and thus the volatility process from tick-by-tick data. Using the resulting volatility estimator on the high frequency data we then illustrate a number of empirical properties of the volatility process. In particular we firstly find a remarkably stable unifractal scaling of the log-increments which was also documented in (Gatheral et al. 2017). We will then consider the distribution of the log-increments, the local properties of the process as measured via the concept of roughness as well as the long term behaviour as measured via persistence. We end the chapter by proposing to model volatility via a fractional Brownian motion and finally consider how roughness has changed historically.

2.1 Estimating Volatility from High-Frequency Data

Let us first set the scene by letting $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ be the filtered probability space under the historical probability measure P . We will assume that the efficient log price of the stock, denoted $\log S_t$, follows a continuous Ito semimartingale process of the form

$$d \log S_t = \mu_t dt + \sigma_t dW_t \quad (2.1)$$

where W_t is a standard Brownian motion adapted to the filtration \mathcal{F}_t and μ_t, σ_t are Cadlag (= "right-continuous with left limits") processes and also adapted to the filtration.¹ In our framework we do not allow the process to contain jumps. We make this fundamental modelling choice since authors such as (Christensen et al. 2014) argues, also using high frequency data, that the proportion of the total variation coming from the jump component on various indices, stocks and exchange rates, including the SPY, is small. We arrive at a similar conclusion which we elaborate on in section 2.2.

2.1.1 Basics

First we wish to somehow obtain historical values of the volatility process σ_t from market data. Unfortunately, this process cannot be directly observed. It is therefore necessary to construct an

¹These assumptions are necessary for the consistency of the integrated variance estimator we will use.

estimator of the process. To this end we wish to review a few well-known concepts.

Following definition 2.2.1 from (Sondermann 2006) the quadratic variation of a real valued continuous function X on $[0, \infty)$ is defined as the function $QV : [0, \infty) \rightarrow \mathbb{R}$ given by²

$$QV(t) = \lim_{n \rightarrow \infty} \sum_{t_i^n \in \tau_n, t_i^n \leq t} (X(t_{i+1}^n) - X(t_i^n))^2 \quad (2.2)$$

where $\tau_n = \{0 = t_0^n < t_1^n < \dots < t_n^n < \infty\}$ is a sequence of increasingly finer partitions satisfying $t_n^n \rightarrow \infty$ and $\sup_{t_i^n \in \tau_n} |t_{i+1}^n - t_i^n| \rightarrow 0$ as $n \rightarrow \infty$. We stress that the definition of course only makes sense if the limit actually exists.

However, since we model $\log S_t$ as an Ito process the quadratic variation $QV(t)$ of $\log S_t$ exists and is exactly equal to the integrated variance $IV(t)$, i.e. we have the relation

$$IV(t) = \int_0^t \sigma_s^2 ds = QV(t). \quad (2.3)$$

Let us now consider how to approximate the volatility process on an interval from t to $t + \Delta$ for some $\Delta > 0$ when we know the integrated variance. A simple approximation would be to consider the volatility process constant in the interval $[t, t + \Delta]$. This is reasonable if the interval is sufficiently small and the process is continuous, i.e. has no jumps.

Defining $IV_t^\Delta := IV(t) - IV(t - \Delta)$ we get the approximation

$$IV_t^\Delta = \int_{t-\Delta}^t \sigma_s^2 ds \approx \Delta \sigma_t^2 \quad (2.4)$$

which implies

$$\sigma_t \approx \sqrt{\frac{IV_t^\Delta}{\Delta}}. \quad (2.5)$$

In practise we don't know the integrated variance either. Thus letting \hat{IV}_t^Δ be some, as of yet unknown, estimator of IV_t^Δ the final estimator of σ_t becomes

$$\hat{\sigma}_t = \sqrt{\frac{\hat{IV}_t^\Delta}{\Delta}}. \quad (2.6)$$

We will occasionally refer to Δ as the sampling frequency of volatility as this interval length will be the finest granularity on which we will have observations of the volatility process. As we have experimented with different choices of sampling frequencies we will generally keep Δ as an abstract symbol.

The question is now how to construct the estimator \hat{IV}_t^Δ from high frequency data. We first consider the case of no jumps and no so-called microstructure noise. Let therefore $Y_t := \log S_t$ be the log-stock price process and assume that we observe the process at the $n + 1$ equidistant time points $(t_i)_{i=0}^n$ in the interval $[t - \Delta, t]$ for some $n \in \mathbb{N}$. That is $t_i = t - \Delta + \frac{\Delta}{n} \cdot i$ for $i = 0, 1, \dots, n$. The realized variance of the log price process over the interval is then defined as

$$RV_t^\Delta(n) := \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^2. \quad (2.7)$$

²When we write t_i^n we don't mean t_i raised to the power n . That is, n is simply the index of the partition τ_n to which t_i^n belongs.

Results from general semi-martingale theory then ensures $RV_t^\Delta(n) \xrightarrow{P} QV_t^\Delta$ as $n \rightarrow \infty$. In the absence of jumps we have $QV_t^\Delta = IV_t^\Delta$ and so also $RV_t^\Delta(n) \xrightarrow{P} IV_t^\Delta$ as $n \rightarrow \infty$. A proof can be found in (Durrett 1996).

In the presence of jumps we still have $RV_t^\Delta(n) \rightarrow QV_t^\Delta$ but now $QV_t^\Delta \neq IV_t^\Delta$ since the quadratic variation also contains a jump component which is then exactly the residual amount $QV_t^\Delta - IV_t^\Delta$. Estimators have been developed to separate out the jump component and the integrated variance. One example is the bi-power variation measure, we will write $BV_t^\Delta(n)$, introduced in (Barndorff-Nielsen & Shephard 2004), which then satisfies $BV_t^\Delta(n) \rightarrow IV_t^\Delta$ even in the presence of jumps. A consistent estimator of the jump component is then given by $RV_t^\Delta(n) - BV_t^\Delta(n)$. In the next section we explain noise robust versions of these two measures which will enable us to later estimate the relative size of the jump component.

2.1.2 Dealing with Microstructure Noise

We consider it well established that when dealing with high frequency data the assumption of the price process being a continuous semi-martingale breaks down. This is so because of so-called microstructure noise. Essentially microstructure noise is what makes the observed prices deviate from the theoretical continuous version of the price process and can with some reason be partly explained by market imperfections such as price discretization, the bid/ask spread and possibly also data errors. An estimator such as (2.7) might thus not be sufficiently robust when sampling on smaller intraday scales.

To formalize the situation we assume that the observed log price process on the interval $[0, T]$ is actually given as

$$Z_{t_i} = Y_{t_i} + \epsilon_{t_i} \quad (2.8)$$

on the observation time points $0 = t_0 < t_1 < \dots < t_n = T$ where ϵ_{t_i} is the microstructure noise at time t_i . As in (Christensen et al. 2014) we will assume the microstructure noise i.i.d. with mean zero, finite fourth moment and also to be independent of the efficient log price process Y .

The question then is how to filter out the effects of the noise process when estimating the integrated variance. There are a number of different approaches one can take to mend this problem. We will estimate the integrated variance using pre-averaging. Many other methods are available in the literature. We choose this one since we find relatively intuitive and easy to implement. Briefly explained, with the pre-averaging method we compute local averages of the returns before squaring the differences. Essentially we are smoothing the process of log returns before estimating the quadratic variation. This smoothing procedure intuitively removes or at least dampens the effect of the microstructure noise. As introduced in (Jacod et al. 2009) the method can be implemented in various ways. We will however follow closely the implementation and mostly also the notation from (Bennedsen et al. 2017a) which refers to (Christensen et al. 2014).

We now formally introduce the method: The goal is to estimate IV_t^Δ . Say we are given observed log tick prices Z_0, Z_1, \dots, Z_n within the interval $[t - \Delta, t]$. The method does not use the actual time points of the observations only that they are in the interval under consideration and are ordered in time. We therefore write Z_i with only the index i to specify the order of the observations.

Choose now an even integer $k_n \geq 2$ potentially depending on the number of observations. The

pre-averaged log returns where we average over a neighbourhood of size k_n are then defined as

$$r_{j,k_n} = \frac{1}{k_n} \left(\sum_{l=k_n/2}^{k_n-1} Z_{j+l} - \sum_{l=0}^{k_n/2-1} Z_{j+l} \right) \quad (2.9)$$

$$j = 0, 1, \dots, n - k_n.$$

As is required for the asymptotic convergence results to hold we need the size of the averaging neighbourhood k_n to scale appropriately with n . Specifically we need $k_n = \theta\sqrt{n} + o(n^{-1/4})$ where θ is a tuning-parameter for the method.

With this in place the following estimators of the integrated variance are proposed:

$$\hat{RV}_t^\Delta = \frac{n}{n - k_n + 1} \frac{1}{k_n \psi_{k_n}} \sum_{j=0}^{n-k_n+1} |r_{j,k_n}|^2 - \frac{\hat{\omega}^2}{\theta^2 \psi_{k_n}} \quad (2.10)$$

$$\hat{BV}_t^\Delta = \frac{n}{n - 2k_n + 2} \frac{1}{k_n \psi_{k_n}} \frac{\pi}{2} \sum_{j=0}^{n-2k_n+1} |r_{j,k_n}| |r_{j+k_n,k_n}| - \frac{\hat{\omega}^2}{\theta^2 \psi_{k_n}} \quad (2.11)$$

Here \hat{RV}_t^Δ is the pre-average version of the realized variance estimator from equation (2.7) and \hat{BV}_t^Δ the equivalent bi-power variation measure. The term $\frac{\hat{\omega}^2}{\theta^2 \psi_{k_n}}$ is a bias correction term which is supposed to compensate for any residual microstructure noise still present after the pre-averaging. Also, in the above equations $\psi_{k_n} = (1 + 2k_n^{-2})/12$ is a constant and $\hat{\omega}^2$ is an estimator of the variance of the microstructure noise, i.e. an estimator of $E\epsilon_{t_1}^2$.

As is done in (Bennedsen et al. 2017a) we use the estimator from (Oomen 2006) which is given by

$$\hat{\omega}^2 = -\frac{1}{n-1} \sum_{j=2}^n r_j r_{j-1} \quad (2.12)$$

where $r_j = Z_j - Z_{j-1}$ is the j 'th observed log return.

Under some general assumptions these estimators are robust to microstructure noise. However only the bi-power-variation estimator \hat{BV}_t^Δ is also robust to jumps in the process. We consider both measures as this allows us to investigate the importance of jumps in estimating the integrated variance on our data sample.

We still need to choose the tuning parameter θ . As we can see θ controls the area over which we average the returns. Choosing a high θ corresponds to averaging over a larger neighbourhood. This would in the extreme case mostly level out the log returns and only leave few observations for the estimators in equation (2.10) and (2.11). Choosing a too small θ would on the other hand not reduce the microstructure induced bias sufficiently. As is proposed in Bennedsen we take $k_n = \lfloor \sqrt{n} \rfloor$ if $\lfloor \sqrt{n} \rfloor$ is even and otherwise $\lfloor \sqrt{n} \rfloor + 1$. This corresponds to a choice of $\theta = 1$. As is claimed by (Bennedsen et al. 2017a) and also shown in (Christensen et al. 2014), the results are not overly sensitive to this choice and we do not investigate this further.

2.1.3 The High Frequency Data

The data we use contains tick-by-tick observations on the SPY on all trading days from January 2006 till the end of 2014 and that from various exchanges, including the NYSE. The ticks are all

recorded during the NYSE trading hours between 09:30 and 16:00 EST.³ We removed holidays and half-open trading days such as the day after thanksgiving to avoid dealing with the low traded volumes around these dates and also excluded a few extra trading days with low traded volumes which caused difficulties when estimating volatility in some small intraday intervals. In total this leaves us with a sample of 2,239 trading days with tick-by-tick data. The average number of trades per day is 249,959.

2.1.4 The Jump Component

In the introduction to this chapter we assumed without further analysis that the stock price contained no jump component. Using the realized variance and bi-power variation measures from equations (2.10) and (2.11) we can now explicitly test this assumption. Letting \hat{RV}_t and \hat{BV}_t respectively be the total realized variance and total bi-power variation incurred up till time t , we choose to estimate the jump component as the difference between \hat{RV}_t and \hat{BV}_t . As already noted this is a consistent estimator of the jump component of the quadratic variation. Sampling each 5 minutes we estimate the jump component to be at 2.6 % of the total quadratic variation, the latter as estimated by \hat{RV}_t . This result is close to the estimate given by (Christensen et al. 2014) in which using high-frequency data he estimates the jump component for various indices, exchange rates and stocks to be no more than a few percent. We conclude that if there are jumps then they are not large enough and/or not of a significant enough intensity to really matter for the estimation of volatility. We will therefore leave the bi-power variation measure behind here and only use the pre-average version of the realized variance estimator going forward.

2.1.5 Filtering Out Intraday Seasonality

It is well known that volatility has a deterministic seasonality component viewed across the day, see for instance (Andersen & Bollerslev 1997). In particular it is found that volatility tends to be higher in the first few trading hours, getting lower around lunch time only to increase somewhat during the afternoon. Many reasons can be proposed for this either U- or J-shape of intraday volatility: The arrival of macroeconomic news tends to happen in the morning and afternoon hours, traders have an extra motivation to rebalance their positions in the early morning hours after receiving knowledge of the overnight news and again at the last opportunity before the market closes at 16:00.

As we wish to consider the volatility process at intraday scales such deterministic calendar effects could influence our results. The goal here is to *deseasonalise* the volatility process to avoid catching unwanted deterministic effects when we want to investigate the stochastic ones and also to hopefully arrive at a stationary version of the volatility process. We will adopt the same basic framework as in (Bennedsen et al. 2017a) and assume the variance process can be decomposed as

$$\sigma_t^2 = \Lambda(t) \cdot \tilde{\sigma}_t^2 \quad (2.13)$$

where $\Lambda(t)$ is a deterministic intraday seasonal factor and $\tilde{\sigma}_t^2$ the deseasonalised process we are really interested in. We will take a simple approach and assume

$$\Lambda(t) = \frac{\sigma_t^2}{\tilde{\sigma}_t^2} \quad (2.14)$$

is constant on each sampling interval of length Δ during the day. Say $\Delta = 5$ minutes. Then to estimate the factor in a specific 5 minute interval, say from 09:00 till 09:05 we simply take the average

³More historical data was available but we found traded volume to be significantly lower in the pre-2006 years. Since we wish to estimate volatility on small time scales we need sufficiently many trades in each interval for precise enough estimates. Therefore this choice.

realized variance of this interval across all observed trading days and then divide by the average 5 minute realized variance across all days and all intraday intervals.

With $\Delta = 5$ minutes this leaves 78 intervals in which to estimate the factor. Using the \hat{RV}_t^Δ estimator we get the following picture:

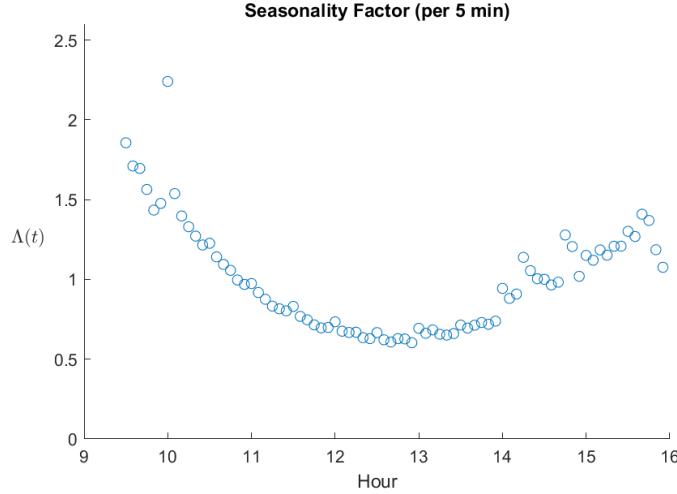


Figure 2.1: Seasonality factor as estimated each 5 minute interval. The data used is the 5 minute sampled realized variance estimates computed using the estimator \hat{RV}_t^Δ .

Here we see that the deterministic intraday variation is quite pronounced and that the commonly observed U or J-shape is clearly visible. Interestingly there is a spike at 10:00 AM. This is most likely due to 10:00 AM being a common time to announce and release macroeconomic news.

With the above our estimates of the deseasonalised variance process $\tilde{\sigma}_t^2$ now becomes

$$\tilde{\sigma}_t^2 = \frac{\hat{IV}_t^\Delta}{\Delta \cdot \Lambda(t)}. \quad (2.15)$$

In the rest of the chapter we will only consider the deseasonalised variance process $\tilde{\sigma}_t^2$. To simplify notation we will refer to it as σ_t^2 even though this is in conflict with the notation used so far.

Finally, while the above deseasonalisation procedure does a lot to remove unwanted seasonal effects we still find a decent amount of seasonal effects in the autocorrelation function of the log volatility process. As an example the orange curve in figure 2.2 shows the autocorrelation function of the log-volatility process as estimated using 5 minute samples and that after the above deseasonalisation procedure. The red curve shows the autocorrelation function before any filtering. Clearly there is still some seasonality present in the data. As we think that the leftover seasonal effects will have an impact on the rest of our analysis we decided to further investigate this feature of the data.

One likely reason why the above procedure is not sufficient to remove all calendar effects is that the seasonality factor is not constant. In fact we find evidence that it has changed moderately between 2006 and 2014 - see figure A.1 in the appendix. If we recompute the seasonality factor for each year in the dataset and use those sets of factors separately for each year we find that the seasonality in the autocorrelation becomes much less pronounced - see the green curve. This result is similar

to what was discovered in (Laakkonen 2007) on the USD/EUR exchange rate. Nonetheless, even recomputing the weights each year does not seem to be sufficient. We therefore settle to recompute them each month. As can be seen from the purple curve we find no visible leftover seasonal effects in the autocorrelation function after doing this.

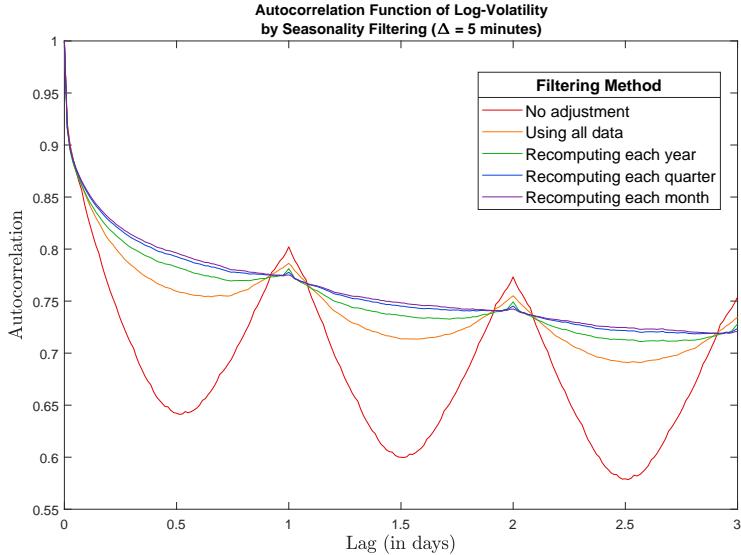


Figure 2.2: The autocorrelation function of log-volatility before and after adjusting for seasonality. The latter is done using weights recomputed at various intervals. The log-volatility process has been sampled each 5 minutes using the realized variance estimator from (2.10).

Using the deseasonalised 5 minute volatility estimates we then tested both the volatility and log-volatility process for a unit-root. Using both an augmented Dickey Fuller test and a Phillips Perron test we reject the hypothesis of a unit-root at p -values below 10^{-3} in all cases. Going forward we will therefore assume that the deseasonalised volatility and log-volatility processes are stationary.

2.2 Empirical Properties of Volatility

In this section we assume the observability of the volatility process σ_t at various sampling frequencies. In practise we use the deseasonalised estimates explained in the previous section. We will not report results for all sampling frequencies to avoid overloading the reader by repeating similar plots and conclusions. We did however experiment with sampling frequencies ranging all the way from 5 minutes to 1 day.⁴

Using the 5 minute sampled volatility estimates obtained from the \hat{RV}_t^Δ estimator we have computed daily volatility estimates. These are shown in annualized terms in figure 2.3 together with the daily last prices. At a first view, we see that the market seems to alternate between periods of high volatility and low volatility. Also, periods of high volatility seem to come in clusters. We

⁴We decided to not sample more frequently than each 5 minutes since we tend to obtain unreasonable estimates if we sample more often; most likely due to a lack of observations in each interval. We do however find that our noise robust estimator works well at the frequencies we choose to consider. Specifically we find that the average realized variance does not change much across the different sampling frequencies ranging from 5 minutes to 1 day. This is an indication that the estimator successfully removes microstructure noise even when sampling each 5 minutes. See section A.2 in the appendix for the results.

furthermore see that the process tends to mean revert back to a normal level after spikes of volatility. Another well-known empirical fact that can be observed in the figure is the so-called leverage effect which says that the stock price and its volatility tends to be negatively correlated. This is even better illustrated in the scatter plot to the right in the figure. A joint modelling of the volatility process and the stock price should thus take this very visible correlation into account.

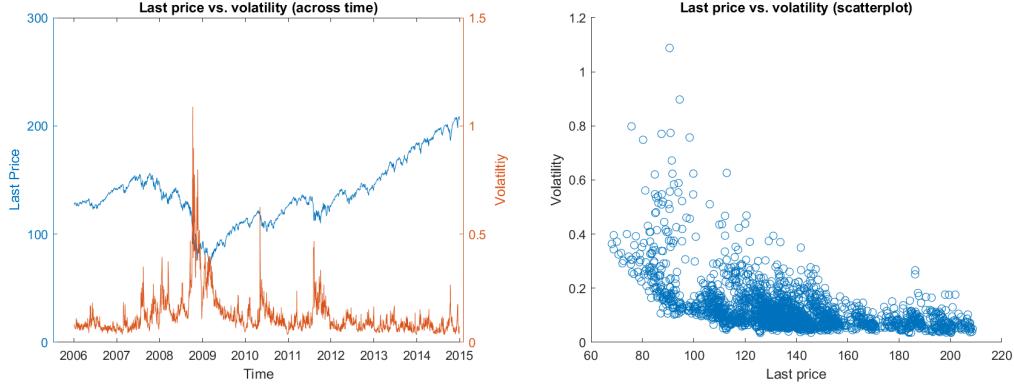


Figure 2.3: Relation between last price and volatility on the SPY. Volatility is calculated using daily realized variance estimates obtained using the estimator from (2.10) by sampling each 5 minutes. Volatility is annualized assuming 252 trading days per year.

2.2.1 Unifractal Scaling

What we will consider in the following are the increments of the log-volatility process, i.e. increments of the form $\log(\sigma_{t+h}) - \log(\sigma_t)$ for $h > 0$. In particular we will investigate what happens when we vary h . As we will see, the increments exhibits some stable scaling properties that will be useful in finding a good model for the volatility process.

Consider therefore the time interval $[0, T]$ over which we observe the volatility process and say we observe it at equidistant time points with time distance Δ . Choose now a time increment $h = K \cdot \Delta$ as some multiple $K \in \mathbb{N}$ of the sampling frequency Δ and set $N = \lfloor \frac{T}{h} \rfloor$. Define now the quantity

$$m(q, h) = \frac{1}{N} \sum_{k=1}^N |\log(\sigma_{kh}) - \log(\sigma_{(k-1)h})|^q \quad (2.16)$$

for any $q \geq 0$.

Assuming the log-volatility process is sufficiently stationary such that some law of large numbers apply, we should expect the above to be an estimator of the quantity⁵

$$E|\log(\sigma_h) - \log(\sigma_0)|^q. \quad (2.17)$$

Computing $m(q, h)$ for different fixed choices of q and across different time increments h we get the following picture:

⁵Actually viewing (2.16) as an estimator of (2.17) we would throw away a large amount of observations if h , equivalently K , is large. In practise we take the average over all such quantities $m(q, h)$ but using different initial time points $\sigma_0, \sigma_\Delta, \sigma_{2\Delta}, \dots, \sigma_{h-\Delta}$. This is equivalent to what is done in (Gatheral et al. 2017).

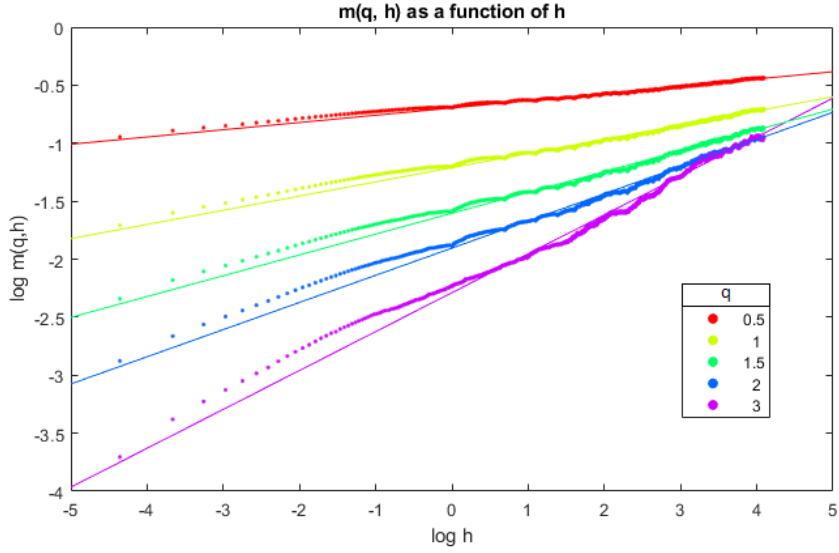


Figure 2.4: Scaling property of the q -moments of the log-volatility increments. Volatility estimates are sampled each 5 minutes. Note that $h = 1$ corresponds to one day in the above plot.

In figure 2.4 we have also added an OLS regression line to each group of data points per choice of q . We see that there appears to be a remarkable stable and linear relationship between $\log h$ and $\log m(q, h)$. Specifically this suggests

$$m(q, h) = K_q h^{\zeta_q} \quad (2.18)$$

where K_q, ζ_q are constants depending on the choice of q . Specifically $\log K_q$ is the intercept seen in the plot and ζ_q the slope.

The above relationship can be related to the concept of multiscaling and the definition of the generalized Hurst exponent. We use the definitions as given in (Matteo 2007).⁶

Definition 2.2.1. *A stochastic process X is multi-scaling if it has stationary increments and satisfies*

$$E(|X_t|^q) = K_q t^{\zeta_q} \quad (2.19)$$

for all $t \in \mathcal{F}$, $q \in \mathcal{L}$ with \mathcal{F}, \mathcal{L} intervals on the real line with $0 \in \mathcal{F}$, $[0, 1] \subset \mathcal{L}$ and K_q and ζ_q functions of q , both with domain \mathcal{L} . We refer to ζ_q as the scaling function.

Fixing t and considering $\log(\sigma_{t+h}) - \log(\sigma_t)$ as a stochastic process with time index $h \geq 0$ we see that our result fits the definition above. In conclusion we have found evidence that log-volatility-increments are multi-scaling.

An important consideration in further investigating the properties of the process is looking at how the quantity ζ_q scales with q . For this purpose it is useful to define the generalized Hurst exponent: Let therefore X be some stochastic process and assume

$$E|X_{t+h} - X_t|^q \propto h^{qH(q)} \quad (2.20)$$

⁶As T. Di Matteo remarks in his paper many concepts in the literature on multi-scaling are often confused. We have ourselves at many times found it difficult to find universally accepted definitions of the relevant concepts. It is therefore possible that the definitions we use here may be defined differently in other works.

holds as a function of h . Then we define the function $H(q)$ as the generalized Hurst exponent.

The definition immediately identifies two classes of processes: 1) Processes where $H(q)$ is constant are called unifractal or uniscaling , 2) processes with $H(q)$ dependent on q are called multifractal or multiscaling. In case (1) we will simplify notation and just write $H(q) = H$ where H is the fixed constant value.

We now wish to investigate to which category the log-volatility process belongs and what $H(q)$ looks like. To this end we look closer at how the slope of the above regressed lines, i.e. ζ_q , depends on q . Hopefully this will reveal if the log-volatility process is uni- or multiscaling.

We get the following picture, where we have also added an OLS regression line:

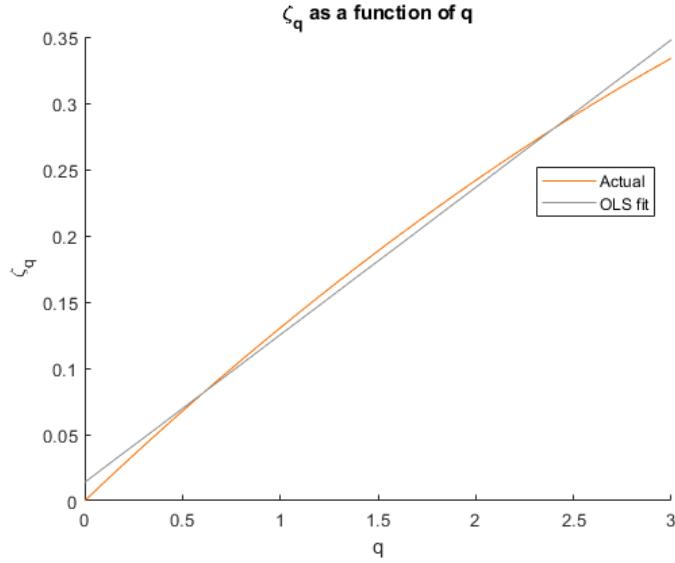


Figure 2.5: The slope ζ_q of the lines from figure 2.4 as a function of q .

While the actual curve appears mildly concave it is still very close to a linear relationship. This suggests that log-volatility is unifractal and that we model

$$\zeta_q = Hq \quad (2.21)$$

where H is the slope of the regression line above. The above results are similar to what was found in (Gatheral et al. 2017) using daily realized volatility estimates. However, here we consider intraday data and confirm the property also holds with data on this time scale. Further, from the above plot we estimate $H = 0.1072$ which is also close to the value obtained in (Gatheral et al. 2017).

The unifractal scaling property suggests one possible modelling choice of the log-volatility increments, namely using self-similar processes. Recall the definition below:

Definition 2.2.2. A stochastic process X is self-similar with scaling exponent H if

$$(X_{ct_1}, \dots, X_{ct_k}) \stackrel{d}{=} (c^H X_{t_1}, \dots, c^H X_{t_k}) \quad (2.22)$$

for all choices of $k \in \mathbb{N}$ and $c, t_1, \dots, t_k \geq 0$.

The reason why self-similar processes are a possible modelling choice is that they are unifractal:

Theorem 2.2.1. *A self-similar process is multi-scaling with a linear scaling function, i.e. it is unifractal.*

Proof. Let X be self-similar stochastic process with scaling exponent H then $X_t \stackrel{d}{=} t^H X_1$. This implies $E(|X_t|^q) = t^{qH} E(|X_1|^q)$. The multi-scaling property therefore holds with $K_q = E(|X_1|^q)$ and $\zeta_q = qH$ and we conclude X is unifractal. \square

From the proof we also see that the scaling exponent H of the self-similar process exactly corresponds to its generalized Hurst exponent.

2.2.2 Distribution

In this subsection we investigate the distribution of the log-increments. In figure 2.6 we show QQ-plots of the increments $\log(\sigma_{t+\Delta}) - \log(\sigma_t)$ for various choices of sampling frequencies Δ .⁷ Here we see that the increments generally are more heavy tailed than a normal distribution. However, if we sample infrequently enough, once per day that is, the normal distribution actually appears as a remarkably good approximation. We do remark that by sampling as infrequently as once per day we really don't think we are sampling the true continuous version of the volatility process σ_t . Rather we are then sampling an aggregated version of it. More specifically sampling infrequently will introduce a bias in the estimation of σ_t via the approximation from equation (2.4). It is conceivable that it is this constant approximation across long time steps Δ that removes some of the non-Gaussianity.

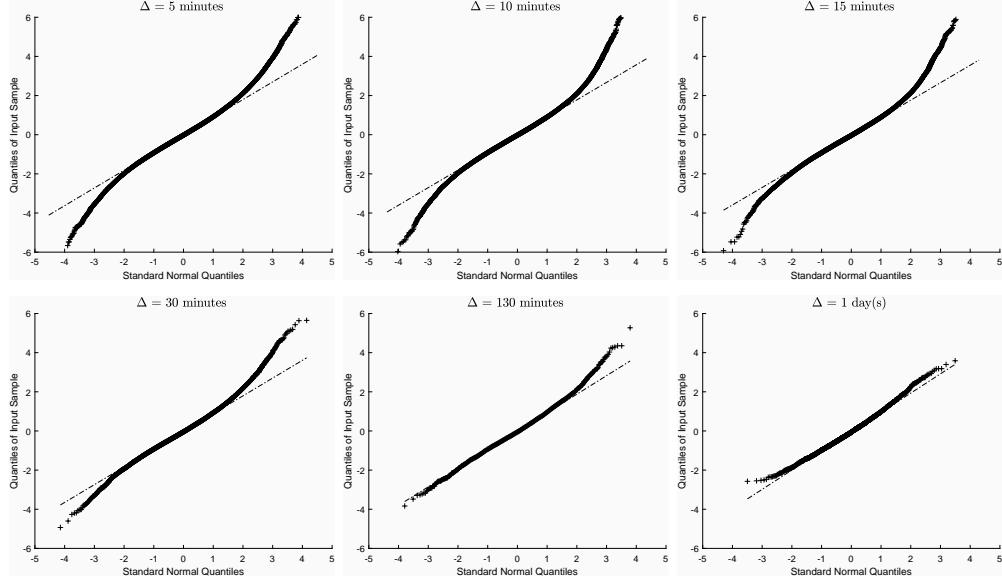


Figure 2.6: QQ-plots of normalised log-volatility increments at various sampling frequencies.

While we will settle for the Gaussian approximation simply to arrive at a tractable model we did also investigate the distribution of log-increments for a fixed small sampling frequency of $\Delta = 5$ minutes and then across lags $h = \Delta \cdot K$ for $K \in \mathbb{N}$. We leave the results to the appendix but remark that we then find the increments to be much more heavy tailed than a normal distribution even for lags of 1 day. In fact only if we consider lags $h = 125$ days do we find the normal approximation

⁷We have normalized all the observations so the means are 0 and the standard deviations 1. This allows us to better compare the plots across the different sampling frequencies.

to be reasonable. In conclusion we therefore think that the actual continuous time version of the log volatility process is non-Gaussian at most relevant lags and that the aggregational Gaussian property, which has also been reported in (Gatheral et al. 2017), is likely to be an artefact of the constant approximation of volatility across long intervals.⁸

2.2.3 Roughness

In this subsection we will investigate the so-called roughness of the log-volatility process. The concept of roughness essentially deals with how the autocorrelation function behaves for infinitely small lags. The measure we are looking for is the fractal index α of our process $\log(\sigma_t)$ which we define if the following relation holds

$$1 - \rho(h) := 1 - \text{Corr}(\log(\sigma_t), \log(\sigma_{t+h})) \sim c|h|^{2\alpha+1} \quad (2.23)$$

in the limit $h \rightarrow 0$ for some constants $c > 0$ and $\alpha \in (-\frac{1}{2}, \infty)$. Here we also defined ρ as the autocorrelation function of $\log \sigma_t$. Also by the notation " \sim " we simply mean that the ratio of the left- and right-hand side tends to 1. We say the process is rough if $\alpha \in (-\frac{1}{2}, 0)$.

Before presenting our estimations of α we wish to provide a little intuition on what it even measures. Recall thus the definition of Hölder continuity:

Definition 2.2.3. *A real valued function f on a d -dimensional Euclidean space is Hölder continuous of order $\phi > 0$ if there exists $C > 0$ s.t. for all s, t in the domain of f we have*

$$|f(t) - f(s)| \leq C|t - s|^\phi. \quad (2.24)$$

We then say a process X is almost surely (a.s.) Hölder continuous of order ϕ if the trajectories of X are a.s. Hölder continuous of order ϕ . Intuitively larger values of the exponent ϕ means the path is more continuous, vice versa.

According to (Bennedsen 2016) it holds that a Gaussian process with stationary increments and fractal index $\alpha \in (-\frac{1}{2}, \frac{1}{2})$ will have a modification with Hölder continuous trajectories of all orders $\phi \in (0, \alpha + \frac{1}{2})$. That is, a lower fractal index means the process is less continuous in the Hölder sense. In particular a standard Brownian motion is Hölder continuous of all orders $\phi \in (0, \frac{1}{2})$. A negative fractal index α thus corresponds to a process that is more rough than a standard Brownian motion. This is one sense in which the fractal index captures the notion of roughness.

The fractal index is, again for Gaussian processes, also connected to the so-called Hausdorff dimension defined below:

Definition 2.2.4. *Consider a pointset $A \subset \mathbb{R}^d$, $d \in \mathbb{N}$, and let $\epsilon > 0$. An ϵ cover of A is a countable collection of balls $\{B_i : i = 1, 2, \dots\}$ in \mathbb{R}^d all with diameter $|B_i| \leq \epsilon$ such that A is contained in the union of the covering set of balls. Let now*

$$H^\delta(A) = \liminf_{\epsilon \rightarrow 0} \left\{ \sum_{i=1}^{\infty} |B_i|^\delta : \{B_i : i = 1, 2, \dots\} \text{ is an } \epsilon \text{ cover of } A \right\} \quad (2.25)$$

be the δ -dimensional Haussdorff measure of A . We finally define the Haussdorff dimension of A as the unique non-negative number D such that $H^\delta(A) = \infty$ for all $\delta < D$ and $H^\delta(A) = 0$ for all $\delta > D$.

⁸In fact (Bennedsen et al. 2017a) shows that a Normal Inverse Gaussian distribution might be more appropriate for this data.

Intuitively D reflects how much space an object occupies and, again intuitively, we should expect rough surfaces to cover more space and thus have a higher Hausdorff dimension. Specifically we may note that a trajectory of a univariate process X defines exactly a point set $\{(t, X_t) : t \geq 0\} \subset \mathbb{R}^2$ and so it makes sense to talk about the Haussdorff dimension of its trajectory. Interestingly, if X is Gaussian with stationary increments then the Haussdorff dimension D of the trajectory is exactly $D = \frac{3}{2} - \alpha$ where α is the fractal index. Thus we see that the fractal index captures the notion of roughness also in the sense of how much space a trajectory covers in \mathbb{R}^2 .

Let us now turn to the question of estimating the fractal index α given our observations of the volatility process. Consider for this purpose the variogram $\gamma_2(h)$ for the process $\log \sigma_t$ defined as

$$\gamma_2(h) = E|\log \sigma_{t+h} - \log \sigma_t|^2. \quad (2.26)$$

Luckily, for covariance-stationary processes the autocorrelation function and variogram are nicely connected as

$$\gamma_2(h) = 2\text{Var}(\log \sigma_t)(1 - \rho(h)). \quad (2.27)$$

Thus if the relation (2.23) holds then also

$$\gamma_2(h) \sim c|h|^{2\alpha+1}. \quad (2.28)$$

Taking logarithm on both sides this implies

$$\log(\gamma_2(h)) \sim \log(c) + (2\alpha + 1)\log(|h|). \quad (2.29)$$

for small h .

This expression allows us to use OLS regression to estimate the slope coefficient $a := (2\alpha + 1)$ and then translate that into an estimate of the fractal index as

$$\hat{\alpha} = \frac{\hat{a} - 1}{2} \quad (2.30)$$

where \hat{a} is the OLS estimate of a .

As the asymptotic approximation is only good for small h it is important that we regress only on observations of small lags. In particular we will only use the points

$\{(\log h, \log \gamma_2(h)) : h = \Delta, 2\Delta, \dots, m\Delta\}$ with $m \in \mathbb{N}$ small. The choice of m , however, introduces a classical bias-variance trade-off as low m 's implies fewer observations for the estimation but a high m introduces bias. We follow the suggestion in (Bennedsen et al. 2017a) and choose $m = 6$. This choice seems to work well for volatilities sampled at low frequencies.

The results for $\Delta = 5$ minutes and $\Delta = 10$ minutes are shown in figure 2.7. With a sampling frequency of 5 minutes we get $\hat{\alpha} = -0.37$ and sampling every 10 minutes we get $\hat{\alpha} = -0.32$. We confirm that volatility indeed is very rough.

We do remark that our estimates of roughness are somewhat dependent on the sampling frequency. As a further example, if we sample instead every 15 minutes we get an estimate of $\hat{\alpha} = -0.29$. Perhaps unsurprisingly it thus seems the process becomes more smooth if we aggregate it across longer time-intervals. In terms of estimating roughness of the actual continuous, or non-aggregated, version of the volatility process we therefore have the most trust in the estimates obtained using $\Delta = 5$ minutes. It is nonetheless possible that the process will turn out to be rougher if we could sample it reliably at even lower frequencies.

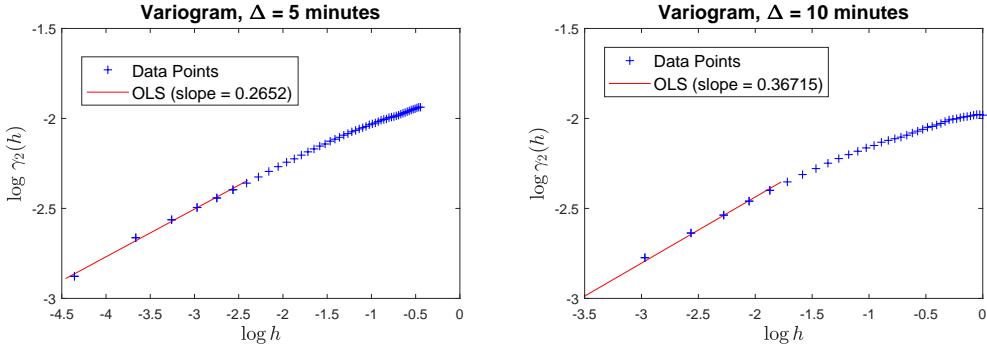


Figure 2.7: OLS regression using the variogram to find the fractal index of the log volatility process. Here $m = 6$. For $\Delta = 5$ minutes we get $\hat{\alpha} = -0.37$ and for $\Delta = 10$ minutes we get $\hat{\alpha} = -0.32$. Note that $h = 1$ corresponds to one day in the plots.

2.2.4 Persistence and Long Memory

In this subsection we will consider the autocorrelations for long time steps. That is, we will investigate the persistence in the autocorrelation function and specifically whether or not volatility has so-called *long memory*. Assume first that the autocorrelation function behaves polynomially at long lags, that is

$$\rho(h) \sim c|h|^{-\beta} \quad (2.31)$$

as $h \rightarrow \infty$ for some $c, \beta > 0$.

We say a stationary process has long range dependence, or long memory, if

$$\int_0^\infty \rho(h) dh = \infty. \quad (2.32)$$

For the case of an (asymptotically) polynomially decaying autocorrelation function we get long memory if $\beta \in (0, 1)$.

Taking log in (2.31) we get the relation

$$\log \rho(h) \sim a + b \log(h) \quad (2.33)$$

for h large enough. Here $a = \log(c)$ and $b = -\beta$. Letting $M', M \in \mathbb{N}$, $M' > M$ and considering the above relation for $h = M\Delta, \dots, M'\Delta$ we can use OLS regression to estimate β . It is important that we choose M, M' large enough for the supposed asymptotics to hold. Nevertheless we limit ourselves to lags below 60 days. We make this choice to ensure that the empirical autocorrelations we use have been estimated using sufficiently many observations. Plotting $\log \rho(h)$ against $\log(h)$ we find by visual inspection that it looks linear if we use all lags from $h = 7$ days to $h = 60$ days. This is true across all sampling frequencies.⁹ We therefore choose M, M' such that $M\Delta = 7$ days and $M'\Delta = 60$ days. We plot the regression for a few sampling frequencies in figure 2.8. In all tested cases we get estimates in the range $\hat{\beta} \in (0.24, 0.26)$. This is somewhat close to the estimates obtained in (Bennedsen et al. 2017a) and clearly suggests long memory.

⁹The reader can assess this him- or herself by glancing at figure A.5 in the appendix.

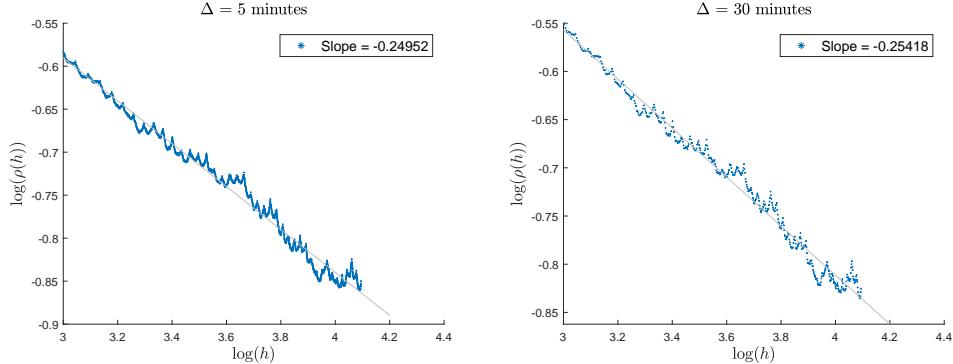


Figure 2.8: OLS regression using the autocorrelation function to find the persistence parameter β . Here $h = M\Delta, \dots, M'\Delta$ with M, M' such that $M\Delta = 7$ days and $M'\Delta = 60$ days. Note that $h = 1$ corresponds to one day in the plots.

2.2.5 Modelling Volatility using a Fractional Brownian Motion

We can summarize our findings so far as follows: The properly deseasonalised log-volatility process has stationary increments, is unifractal, rough, more heavy tailed than a normal distribution and has long memory. In spite of these detailed findings we will now make a few simplifications to arrive at a more tractable model. Firstly, from the unifractal scaling relationship it is natural to assume that log-volatility is self-similar. If we further assume it Gaussian, which is reasonable if we only sample it daily, we get a unique process (up to location and scale); namely the fractional Brownian motion (fBM):

Definition 2.2.5. *The fractional Brownian motion $(W_t^H)_{t \geq 0}$ is characterized by the following three properties:*

- (1) *the process is Gaussian with zero mean*
- (2) *it has stationary increments*
- (3) *it is self-similar with index $H \in (0, 1)$.*

As is usual we also require $W_0^H = 0$.

Part (3) in the definition can equivalently be replaced by requiring

$$E(W_t^H W_s^H) = \frac{1}{2} (|t|^{2H} + |s|^{2H} + |t - s|^{2H}). \quad (2.34)$$

As the process has Gaussian increments with a known covariance structure we can easily simulate the process on a given time grid using for instance Cholesky factorisation. Doing this we in figure 2.9 plot trajectories of the fractional Brownian motion for $H = 0.1, 0.5, 0.9$. Here we see experimentally that lower values of H is associated with more rough sample paths. Also note that for $H = 0.5$ we get back the standard Brownian motion as then $E(W_t^{0.5} W_s^{0.5}) = |t - s|$.

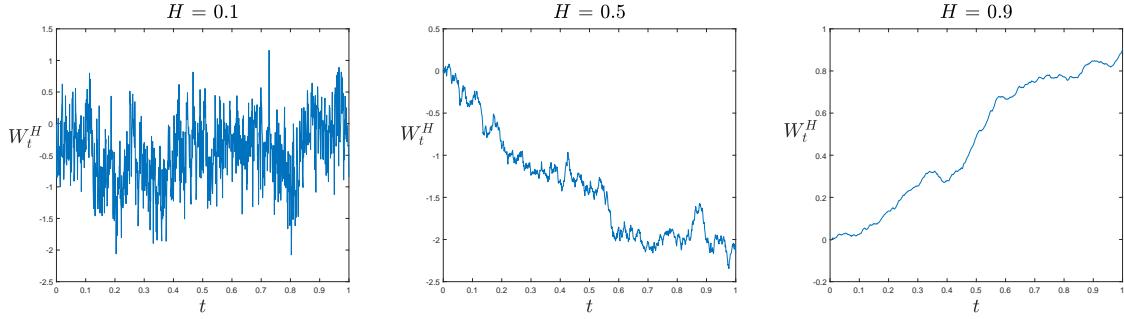


Figure 2.9: Simulated fractional Brownian motion paths for different values of H . Paths are simulated on 1.000 equidistant time points in the unit interval using Cholesky factorization.

Let us now consider theoretically how well the fBM captures roughness and long memory. Using the covariance requirement (2.34) one can show that a fBM has long memory parameter $\beta = 2(1-H)$ and fractal index $\alpha = H - \frac{1}{2}$ where H is the self-similarity index as given in definition 2.2.5. Thus for a fBM long memory and roughness are intimately connected via H . The relation $\alpha = H - \frac{1}{2}$ also clearly explains why the sample paths in figure 2.9 look more rough for low values of H . The relations unfortunately implies that to have long memory one has to sacrifice roughness and vice versa. Thus we will have to choose. The fractional stochastic volatility (FSV) model proposed in (Comte & Renault 1998) uses a fBM with $H > \frac{1}{2}$ to exhibit long memory whereas the Rough FSV (RFSV) model proposed in (Gatheral et al. 2017) uses $H < \frac{1}{2}$ to get roughness. We will follow the latter approach and assume $H < \frac{1}{2}$ since we find this leads to a class of models that work well for option pricing.

Nonetheless, since log-volatility at high frequency time scales is not Gaussian the relationship $\alpha = H - \frac{1}{2}$ cannot generally be assumed to hold. This raises the question how one should compute the H parameter of the fBM in our volatility model. For instance, from the 5 minute samples and using the unifractal scaling relationship from figure 2.2.1 we found $H = 0.11$. Computing it instead via the fractal index we get $\hat{\alpha} + \frac{1}{2} = -0.37 + \frac{1}{2} = 0.13$. While these values appear somewhat similar the difference may still have large implications when we consider option pricing under the model. There is thus an important question in which method is the most reasonable to obtain H in our RFSV model. We remark that there is a third method since, as we will see, H is preserved under a general change of measure and so a market implied version of it can also be extracted from options data. It is however still useful to also estimate it under the historical probability measure. This is particularly so if one wants to make predictions of H for the purpose of betting on it in a misspecified market.

As an extension of these thoughts let us consider if and how H in our RFSV model varies across time. Sampling volatility every 5 minutes we have estimated the fractal index α as in section 2.2.3 with $m = 6$ in rolling non-overlapping windows of 10 trading days. With this we show our estimates of $\alpha + \frac{1}{2}$ across time in figure 2.10.

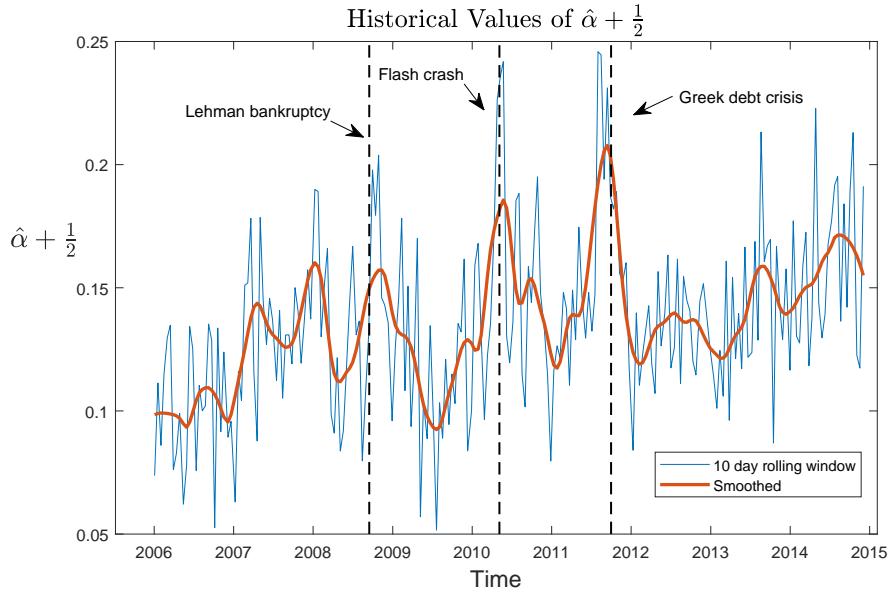


Figure 2.10: Historical estimates of $\hat{\alpha} + \frac{1}{2}$ where the fractal index α is estimated using the variogram method from section 2.2.3 with $m = 6$. Estimates are computed in non-overlapping intervals with a length of 10 trading days and that using the 5 minute sampled volatility estimates.

What we see is that roughness, and by extension H , does not appear to be very constant across time. In fact the overall level seems to have changed from being very rough in 2006 to being a little less rough in 2014. The values also seem to fluctuate a lot. At first one might conjecture that this simply is a result of noise in the estimates due to a finite sample size in each window. Our own guess is however that many of the wild fluctuations are actual reflections of changes in roughness across short periods of time. We make this claim since many of the peaks appear to be connected to actual significant financial events. In particular we have marked out the Lehman Brothers bankruptcy which was filed on the 15th of September 2008, the Flash Crash which occurred on the 6th of May 2010 and finally the Greek debt crisis which we have marked at October 1st 2011 that being a date we think is around the period when the crisis peaked.

While a time-varying level of roughness goes slightly against using a fBM to model volatility since it has a constant level of roughness, it also opens the door to placing bets on what the future value of roughness (or H) will be. In relation to this it would be interesting to more closely characterise what causes roughness in volatility or atleast what it correlates with. While we leave that for future research we will in this thesis instead provide some initial thoughts on how one can even realize a PnL against a market that uses an incorrect H . More about this in chapter 6.

Chapter 3

Pricing Model

The previous chapter suggests that we model the log-volatility increments as

$$\log(\sigma_u) - \log(\sigma_t) = \nu (W_u^H - W_t^H) \quad (3.1)$$

for all $u > t \geq 0$ where $\nu > 0$ is a constant and $(W_t^H)_{t \geq 0}$ a fractional Brownian motion with self-similarity index $H < \frac{1}{2}$.

The goal of this chapter is to apply a change of measure to go from the model under the physical probability measure P to a model under the risk-neutral probability measure Q . This will allow us to price options and other derivatives under the model.

3.1 Model Under P

As it is more convenient, we will in the rest of this chapter consider the variance process $v_u := \sigma_u^2$ instead of the volatility process σ_u . Also, before attempting to apply a change of measure it is convenient to rewrite the representation of the volatility (or variance) process a little. Consider for this purpose the Mandelbrot-Van Ness representation of a fractional Brownian motion

$$W_t^H = C_H \left(\int_{-\infty}^t \frac{dW_s^P}{(t-s)^\gamma} - \int_{-\infty}^0 \frac{dW_s^P}{(-s)^\gamma} \right) \quad (3.2)$$

for $t \geq 0$. Here $(W_s^P)_{s \in \mathbb{R}}$ is a two-sided standard Brownian motion¹ under the observable probability measure P and $\gamma = \frac{1}{2} - H$ and $C_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}$ constants.

Inserting this into (3.1) and writing it in terms of the variance process v_u we get the following:

$$\log v_u - \log v_t = 2\nu C_H \left\{ \int_{-\infty}^u \frac{dW_s^P}{(u-s)^\gamma} - \int_{-\infty}^0 \frac{dW_s^P}{(-s)^\gamma} - \left(\int_{-\infty}^t \frac{dW_s^P}{(t-s)^\gamma} - \int_{-\infty}^0 \frac{dW_s^P}{(-s)^\gamma} \right) \right\} \quad (3.3)$$

$$= 2\nu C_H \left\{ \int_{-\infty}^u \frac{dW_s^P}{(u-s)^\gamma} - \int_{-\infty}^t \frac{dW_s^P}{(t-s)^\gamma} \right\} \quad (3.4)$$

$$= 2\nu C_H \left\{ \int_t^u \frac{dW_s^P}{(u-s)^\gamma} + \int_{-\infty}^t \left[\frac{1}{(u-s)^\gamma} - \frac{1}{(t-s)^\gamma} \right] dW_s^P \right\} \quad (3.5)$$

$$= 2\nu C_H \{M_t(u) + Z_t(u)\} \quad (3.6)$$

¹We simply define a two-sided standard Brownian motion as the process $(W_s^P)_{s \in \mathbb{R}}$ s.t. $(W_s^P)_{s \geq 0}$ and $(W_{-s}^P)_{s \geq 0}$ are independent standard Brownian motions.

where we in the last equality have defined

$$M_t(u) := \int_t^u \frac{dW_s^P}{(u-s)^\gamma} \quad (3.7)$$

and

$$Z_t(u) := \int_{-\infty}^t \left[\frac{1}{(u-s)^\gamma} - \frac{1}{(t-s)^\gamma} \right] dW_s^P. \quad (3.8)$$

One may here note that $Z_t(u)$ is \mathcal{F}_t measurable and also that $M_t(u)$ is independent of \mathcal{F}_t and is Gaussian. The mean is clearly zero and using the change-of-variable $l = u - s$ we can explicitly compute the variance as

$$\int_t^u (u-s)^{-2\gamma} ds = \int_0^{u-t} l^{-2\gamma} dl = \left[\frac{1}{1-2\gamma} l^{1-2\gamma} \right]_0^{u-t} = \frac{(u-t)^{2H}}{2H}. \quad (3.9)$$

Define now

$$\tilde{W}_t^P(u) := \sqrt{2H} \cdot M_t(u) = \sqrt{2H} \int_t^u (u-s)^{-\gamma} dW_s^P \quad (3.10)$$

and note that $\tilde{W}_t^P(u)$ will then also be normally distributed with mean zero except the variance is now $(u-t)^{2H}$.

Defining also $\eta := 2\nu C_H / \sqrt{2H}$ and then using (3.3) - (3.6) we get

$$v_u = v_t \exp(2\nu C_H [M_t(u) + Z_t(u)]) = v_t \exp\left(\eta \tilde{W}_t^P(u) + 2\nu C_H Z_t(u)\right) \quad (3.11)$$

which concludes the first part of our rewriting of the variance process. However, annoyingly the above expression depends on $Z_t(u)$ which depends on W_s going all the way back to $s = -\infty$. Our next step is therefore to encapsulate the dependence on $Z_t(u)$ in another more natural object which turns out to be the conditional expectation $E^P(v_u | \mathcal{F}_t)$.

To this end recall first that $Z_t(u) \in \mathcal{F}_t$ and that $\tilde{W}_t^P(u)$ is independent of \mathcal{F}_t and Gaussian with mean zero such that

$$E^P(v_u | \mathcal{F}_t) = v_t \exp\left(2\nu C_H Z_t(u) + \frac{1}{2}\eta^2 E(\tilde{W}_t^P(u)^2)\right). \quad (3.12)$$

As we also wish to write v_u in terms of the so-called stochastic exponential we state a (limited) definition for continuous semi-martingales:

Definition 3.1.1. For a continuous semi-martingale X_t the stochastic exponential process $\mathcal{E}(\cdot)(t)$ is defined as

$$\mathcal{E}(X)(t) = e^{X_t - X_0 - \frac{1}{2}[X]_t} \quad (3.13)$$

where $[X]_t$ is the quadratic variation process.

For a stochastic Brownian integral X with deterministic integrand we then get

$$\mathcal{E}(X)(t) = e^{X_t - \frac{1}{2}E(X_t^2)}. \quad (3.14)$$

The above is the only context in which we will use the stochastic exponential and so we will change notation to $\mathcal{E}(X_t)$ instead of writing $\mathcal{E}(X)(t)$. Also, we will not specify with notation under which probability measure the expectation in (3.14) is to be taken. It should however be clear from the context.

With the above we arrive at our final representation for v_u :

$$v_u = v_t \exp \left(\eta \tilde{W}_t^P(u) + 2\nu C_H Z_t(u) \right) \quad (3.15)$$

$$= v_t \exp \left(2\nu C_H Z_t(u) + \frac{1}{2} \eta^2 E^P(\tilde{W}_t^P(u)^2) \right) \cdot \exp \left(\eta \tilde{W}_t^P(u) - \frac{1}{2} \eta^2 E^P(\tilde{W}_t^P(u)^2) \right) \quad (3.16)$$

$$= E^P(v_u | \mathcal{F}_t) \mathcal{E}(\eta \tilde{W}_t^P(u)). \quad (3.17)$$

Although we are still under P , an important thing to note at this stage is that v_u only depends on the past information through the variance forecast $E^P(v_u | \mathcal{F}_t)$. This property will be carried over when we apply a change of measure except the past dependence will instead be captured in the forward variance curve (to be defined).

3.2 Volatility Process Under Q

To ease the exposition we start by just considering a change of measure applied to the volatility process. Thus assuming that the filtration is the one generated by the underlying Brownian motion W_t and applying a general change of measure via a Girsanov transformation we get

$$dW_s^P = dW_s^Q + \phi_s ds \quad (3.18)$$

where ϕ_s is the Girsanov kernel for the transformation satisfying of course sufficient conditions for the associated likelihood process to be a martingale.²

With this transformation we can write

$$v_u = E^P(v_u | \mathcal{F}_t) \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{1}{(u-s)^\gamma} dW_s^P - \frac{\eta^2}{2} (u-t)^{2H} \right\} \quad (3.19)$$

$$= E^P(v_u | \mathcal{F}_t) \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{1}{(u-s)^\gamma} (dW_s^Q + \phi_s ds) - \frac{\eta^2}{2} (u-t)^{2H} \right\} \quad (3.20)$$

$$= E^P(v_u | \mathcal{F}_t) \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{1}{(u-s)^\gamma} dW_s^Q - \frac{\eta^2}{2} (u-t)^{2H} + \eta \sqrt{2H} \int_t^u \frac{\phi_s}{(u-s)^\gamma} ds \right\} \quad (3.21)$$

$$= E^P(v_u | \mathcal{F}_t) \mathcal{E} \left(\eta \tilde{W}_t^Q(u) \right) \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{\phi_s}{(u-s)^\gamma} ds \right\}. \quad (3.22)$$

Here $\tilde{W}_t^Q(u)$ is simply notation for (3.10) but where the P -Brownian motion is replaced by a Q -Brownian motion. Naturally the stochastic exponential will also have to be understood in the context of a Q -expectation.

Whereas under P the variance process is log-normally distributed we see that it might not be the case under Q depending on the form of ϕ_s . If ϕ_s is deterministic then it will still be log-normal but generally that won't be true. To arrive at a more tractable model we will however make exactly that assumption. I.e. we assume $\phi_s = \phi(s)$ where $\phi(s)$ is a deterministic function of time. The specific model we obtain with this choice is what we call the *rough Bergomi model*.

Define now the forward variance curve $\xi_t(u) := E^Q(v_u | \mathcal{F}_t)$. Under our assumptions it will have

²Satisfying the Novikov condition would be sufficient.

a very explicit form:³

$$\xi_t(u) = E^Q(v_u | \mathcal{F}_t) \quad (3.23)$$

$$= E^Q \left(E^P(v_u | \mathcal{F}_t) \mathcal{E} \left(\eta \tilde{W}_t^Q(u) \right) \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{\phi(s)}{(u-s)^\gamma} ds \right\} \middle| \mathcal{F}_t \right) \quad (3.24)$$

$$= E^P(v_u | \mathcal{F}_t) \underbrace{E^Q \left(\mathcal{E} \left(\eta \tilde{W}_t^Q(u) \right) \middle| \mathcal{F}_t \right)}_{= E^Q(\mathcal{E}(\eta \tilde{W}_t^Q(u))) = 1} \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{\phi(s)}{(u-s)^\gamma} ds \right\} \quad (3.25)$$

$$= E^P(v_u | \mathcal{F}_t) \exp \left\{ \eta \sqrt{2H} \int_t^u \frac{\phi(s)}{(u-s)^\gamma} ds \right\}. \quad (3.26)$$

Combining this with (3.19) - (3.22) we conclude

$$v_u = \xi_t(u) \mathcal{E} \left(\eta \tilde{W}_t^Q(u) \right). \quad (3.27)$$

We stress the beauty of this representation as all dependence on the past is contained in the forward variance curve $\xi_t(u)$. Importantly this object may be observed on the market, either directly or indirectly via option prices, and so for our purposes it is a known function at any given time.

3.3 Full Model Under Q

We will now formally state and then apply a change of measure to the full model. To this end we first need to modify our notation a little as we will now deal with two Brownian motions instead of one. Thus assume we have two independent P -Brownian motions $W_{1,t}^P$ and $W_{\perp,t}^P$ those being the only stochastic factors. Let now $\rho \in [-1, 1]$ and define a new Brownian motion as

$$W_{2,t}^P := \rho W_{1,t}^P + \sqrt{1-\rho^2} W_{\perp,t}^P$$

such that $dW_{1,t}^P dW_{2,t}^P = \rho dt$.

We formally state our model under P as

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_{2,t}^P \quad (3.28)$$

$$v_u = E^P(v_u | \mathcal{F}_t) \mathcal{E} \left(\eta \tilde{W}_t^P(u) \right) \quad (3.29)$$

where $\tilde{W}_t^P(u)$ is defined in terms of the Brownian motion $W_{1,t}^P$. The correlation between the Brownian motions $W_{1,t}^P$ and $W_{2,t}^P$ is then what gives the possibility of a leverage effect between the stock and the volatility. We will furthermore assume the presence of a risk-free asset with a local rate of return of r_t and assume that the stock pays a continuous and proportional dividend yield of q_t . As in the rest of the thesis we will assume $r_t = r(t)$ and $q_t = q(t)$ where $r(t)$ and $q(t)$ are deterministic functions of time.

If we let the filtration \mathcal{F}_t be the one generated by $(W_{1,t}^P, W_{\perp,t}^P)'$ then any measure transformation will come about as a Girsanov transformation with some Girsanov kernel $\phi_t = (\phi_{1,t}, \phi_{\perp,t})'$ for which it then holds that

$$\begin{pmatrix} dW_{1,t}^P \\ dW_{\perp,t}^P \end{pmatrix} = \begin{pmatrix} \phi_{1,t} \\ \phi_{\perp,t} \end{pmatrix} dt + \begin{pmatrix} dW_{1,t}^Q \\ dW_{\perp,t}^Q \end{pmatrix} \quad (3.30)$$

³In the second last equality we use that $E(\mathcal{E}(X)) = E(e^{X - \frac{1}{2}EX^2}) = e^{0 + \frac{1}{2}EX^2 - \frac{1}{2}EX^2} = 1$ for X being Gaussian with mean 0.

where $W_{1,t}^Q$ and $W_{\perp,t}^Q$ are independent Brownian motions under the new probability measure Q .

In order to find an equivalent martingale measure we need to choose the Girsanov kernel such that the drift of the stock under such a measure is exactly $r_t - q_t$. Using a generic change of measure we have that

$$dW_{2,t}^P = \rho dW_{1,t}^P + \sqrt{1-\rho^2} dW_{\perp,t}^P \quad (3.31)$$

$$= \rho (\phi_{1,t} dt + dW_{1,t}^Q) + \sqrt{1-\rho^2} (\phi_{\perp,t} dt + dW_{\perp,t}^Q) \quad (3.32)$$

$$= (\rho \phi_{1,t} + \sqrt{1-\rho^2} \phi_{\perp,t}) dt + \rho dW_{1,t}^Q + \sqrt{1-\rho^2} dW_{\perp,t}^Q \quad (3.33)$$

$$=: \phi_{2,t} dt + dW_{2,t}^Q \quad (3.34)$$

where we have defined $\phi_{2,t}$ and $W_{2,t}^Q$ appropriately.

This implies that

$$dS_t = \mu_t S_t dt + \sigma_t S_t (\phi_{2,t} dt + dW_{2,t}^Q) \quad (3.35)$$

$$= (\mu_t + \sigma_t \phi_{2,t}) S_t dt + \sigma_t S_t dW_{2,t}^Q. \quad (3.36)$$

Thus the requirement becomes

$$\mu_t + \sigma_t \phi_{2,t} = r_t - q_t. \quad (3.37)$$

Let us therefore choose $(\phi_{1,t}, \phi_{\perp,t})'$ such that the above holds almost surely for all $t \geq 0$. Since there is no further requirement on the Girsanov kernel besides that the resulting likelihood process should be a martingale and recalling further that

$$\phi_{2,t} = \rho \phi_{1,t} + \sqrt{1-\rho^2} \phi_{\perp,t} \quad (3.38)$$

we can in effect choose $\phi_{1,t}$ essentially as we want.⁴ Especially we can assume it is a deterministic function of time which allows us to use the results from the last section. Thus assuming $\phi_{1,t} = \phi_1(t)$ with $\phi_1(t)$ a deterministic function we can write the resulting rough Bergomi model under Q as

$$dS_t = (r_t - q_t) S_t dt + \sigma_t S_t dW_{2,t}^Q \quad v_t = \xi_0(t) \mathcal{E}(\eta \tilde{W}_0^Q(t)) \quad (3.39)$$

where $\tilde{W}_0^Q(t)$ is driven by the Brownian motion $W_{1,t}^Q$ and we clearly have that $dW_{1,t}^Q dW_{2,t}^Q = \rho dt$. Note that if we disregard the forward variance curve, which we argue can be observed on the market, the model only contains three parameters: H, η and ρ .

Let us now consider potential ways of estimating option prices in the rough Bergomi model. Unfortunately the model is not Markov in S_t and v_t since $\xi_t(u) = E^Q(v_u | \mathcal{F}_t) \neq E^Q(v_u | v_t)$. To see this recall the expression for $E^Q(v_u | \mathcal{F}_t)$ in equations (3.23) - (3.26) and combine this with the expression for $E^P(v_u | \mathcal{F}_t)$ in (3.12). As $Z_t(u)$ in (3.12) depends on the entire history of W_1^P we get $E^P(v_u | \mathcal{F}_t) \neq E^P(v_u | v_t)$ and so $E^Q(v_u | \mathcal{F}_t) \neq E^Q(v_u | v_t)$. In conclusion there is no pricing function in just the two state variables S_t and v_t and so no simple PDE can be obtained. As can be seen from the representation of the variance process in (3.27) the model is however Markov in S_t and ξ_t where $\xi_t := (\xi_t(u))_{u>t}$ is the entire forward variance curve as observed at time t . However, as we will see in chapter 6 this will only lead to an infinite-dimensional PDE and so is not very practical for computing option prices either. In conclusion we will have to estimate option prices via Monte Carlo.

⁴This is of course not surprising given that we consider a stochastic volatility model with two stochastic factors and only one asset in excess of the risk-free asset.

Simulation and Monte Carlo estimation is exactly what we will consider in the next chapter. Nonetheless, to simplify our exposition there we will now spend a little time reformulating the pricing model in terms of the forward price. The assumptions of a deterministic interest rate and dividend yield are essential here. First let $F_t = e^{\int_t^T r(s) - q(s) ds} S_t$ be the maturity $T > 0$ forward price as observed at time $t \leq T$ where we let the maturity be implicit in our notation for the forward price.

A quick application of Ito under these assumptions will then show that

$$dF_t = \sigma_t F_t dW_{2,t}^Q. \quad (3.40)$$

Consider now a simple claim on the stock with payoff function $g(\cdot)$ and that with expiry at T . The time t price of this claim can then be written as

$$E^Q \left(e^{\int_t^T -(r(s) - q(s)) ds} g(S_T) | \mathcal{F}_t \right) = D(t, T) \cdot E^Q (g(F_T) | \mathcal{F}_t) \quad (3.41)$$

where we have defined $D(t, T) = e^{\int_t^T -(r(s) - q(s)) ds}$ and used that $F_T = S_T$. As the discount factor $D(t, T)$ may be easily computed from the functions $r(s), q(s)$ we are only left with estimating the expectation $E^Q (g(F_T) | \mathcal{F}_t)$. Fortunately recalling (3.40) we may think of $E^Q (g(F_T) | \mathcal{F}_t)$ as the price of a simple claim on a stock in an equivalent economy except the interest rate and dividend yield are both zero and the current value of the stock is F_t .

Thus for the purpose of the Monte Carlo techniques we will present in the next chapter we will without loss of generality assume $r(t) = q(t) = 0$. Also, to simplify notation we will from now on drop the explicit Q -measure dependence in the model and just write $\tilde{W}, W_\perp, W_1, W_2$ etc. It is then implicit that we refer to the Q -model.

Chapter 4

Simulation

In this chapter we consider various methods for simulating the rough Bergomi Model. For convenience we only consider the forward price formulation of the pricing model as presented in the last chapter. More specifically we assume $r(t) = q(t) = 0$ without loss of generality. We will furthermore assume that the forward variance curve $t \mapsto \xi_0(t)$ is known. In chapter 5 we explain how to estimate it from option prices.

The chapter is structured as follows: First we briefly consider a naive Riemann sum approximation to simulating both the stock and variance process. As we find this method produces a significant distributional bias we next consider exact simulation of the variance process. However, it turns out exact simulation is not the fastest method available and since we need a reasonably fast simulation scheme to make calibration feasible we therefore also consider a hybrid scheme as presented in (Bennedsen et al. 2017b). Finally we consider a few variance reduction techniques as proposed in (McCrickerd & Pakkanen 2018).

4.1 A Simple Riemann Scheme

Let us first present the simplest simulation scheme we can think of. Say we want to simulate the model across the interval $[0, T]$, $T > 0$, and that on some equidistant discretization $0, \frac{1}{n}, \frac{2}{n}, \dots, \frac{\lfloor nT \rfloor}{n}$ for some chosen $n \in \mathbb{N}$.¹ Define also $M := \lfloor nT \rfloor$ and $t_i := \frac{i}{n}$ for $i = 0, 1, \dots, M$. A naive simulation scheme will use a Riemann sum approximation for both the stock and variance process. Specifically we approximate

$$\tilde{W}_0(t_i) = \sqrt{2H} \int_0^{t_i} (t_i - s)^{-\gamma} dW_{1,s} \approx \sqrt{2H} \sum_{j=0}^{i-1} (t_i - t_j)^{-\gamma} (W_{1,t_{j+1}} - W_{1,t_j}) \quad (4.1)$$

and furthermore

$$S_{t_i} = S_{t_{i-1}} + \sqrt{v_{t_{i-1}}} S_{t_{i-1}} (W_{2,t_i} - W_{2,t_{i-1}}) \quad (4.2)$$

for $i = 1, 2, \dots, M$.

Simulation can then be performed as follows: For $i = 1, \dots, M$ simulate first the increment $W_{1,t_i} - W_{1,t_{i-1}} \sim \mathcal{N}(0, \frac{1}{n})$. At this stage we have then already simulated the increments $(W_{1,t_{j+1}} - W_{1,t_j})_{j=0}^{i-1}$ and so may compute $\tilde{W}_0(t_i)$ from (4.1). This will then give us the variance process at time t_i from

¹The discretisation technically does not entirely cover $[0, T]$ in general. However $\frac{\lfloor nT \rfloor}{n} \rightarrow T$ for $n \rightarrow \infty$.

the representation

$$v_{t_i} = \xi_0(t_i) \mathcal{E} \left(\eta \tilde{W}_0(t_i) \right). \quad (4.3)$$

Simulating also $W_{\perp,t_i} - W_{\perp,t_{i-1}} \sim \mathcal{N}(0, \frac{1}{n})$ and then computing

$$W_{2,t_i} - W_{2,t_{i-1}} = \rho (W_{1,t_i} - W_{1,t_{i-1}}) + \sqrt{1-\rho^2} (W_{\perp,t_i} - W_{\perp,t_{i-1}}) \quad (4.4)$$

we may then from (4.2) compute the stock price S_{t_i} .

While all the simulation methods we will consider use the Euler scheme in (4.2) for the stock price it turns out that the simple Riemann sum approximation to the variance process produces a noticeable distributional bias. The implications for estimating option prices are significant. As an example we considered an at-the-money option with expiry $T = \frac{1}{2}$ and found that simple Monte Carlo estimation of the option price using even 1.000 steps per year results in a relative bias of roughly 15 percent.² We therefore do not think the method is a relevant candidate for simulating the model. We explain why it does not work when we explain our Hybrid scheme in section 4.3.

4.2 Exact Simulation

In light of the distributional bias produced by the naive Riemann scheme we here consider exact simulation of the variance process or more specifically of $\tilde{W}_0(t)$. Firstly, if we look at the process $\tilde{W}_0(t) = \sqrt{2H} \int_0^t (t-s)^{-\gamma} dW_{1,s}$ we see that the kernel $(t-s)^{-\gamma}$ depends on the upper limit t in the stochastic integral. Because of this we cannot simulate the processes step by step if we want to simulate it exactly. Instead we need to simultaneously simulate the entire vector

$$\left(W_{2,t_1}, \dots, W_{2,t_M}, \tilde{W}_0(t_1), \dots, \tilde{W}_0(t_M) \right)' . \quad (4.5)$$

One will quickly note that the above vector is Gaussian and that with mean zero. Thus all there is left to know is the covariance matrix. In terms of notation we write this matrix as

$$\Sigma = \begin{pmatrix} \Sigma_{W,W} & \Sigma_{W,\tilde{W}} \\ \Sigma_{\tilde{W},W} & \Sigma_{\tilde{W},\tilde{W}} \end{pmatrix} \quad (4.6)$$

where each block matrix, e.g. $\Sigma_{W,W}$, is $M \times M$. Thus the entire covariance matrix is $2M \times 2M$. Each entry can be calculated symbolically and while we will present those results below we delegate the derivations to the appendix.

Let $u, v > 0$. Then it can be shown that

$$\text{Cov}(W_{2,u}, W_{2,v}) = \min(u, v). \quad (4.7)$$

It also holds that

$$\text{Cov}\left(W_{2,u}, \tilde{W}_0(v)\right) = \rho D_H \left(v^{H+1/2} - (v - \min(u, v))^{H+1/2} \right) \quad (4.8)$$

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

Finally, for $v > u$ we get the following covariances

$$\text{Cov}\left(\tilde{W}_0(u), \tilde{W}_0(v)\right) = u^{2H} G\left(\frac{u}{v}\right) \quad (4.9)$$

²This and further results highlighting the distributional bias are shown and explained in section A.6 in the appendix.

with

$$G(x) = 2H \int_0^1 \frac{1}{(1-s)^\gamma (x-s)^\gamma} ds = \frac{1-2\gamma}{1-\gamma} x^\gamma {}_2F_1(1, \gamma, 2-\gamma, x) \quad (4.10)$$

where $x \leq 1$ and where ${}_2F_1(\cdot)$ is the so-called confluent hypergeometric function.

The recipe for simulating the stock is therefore as follows: First compute the covariance matrix Σ using the above formulas. Then simulate the entire vector in (4.5) from the normal distribution with mean zero and covariance matrix Σ .³ The variance process is now easily computed from equation (4.3) and the Euler scheme in (4.2) gives us the stock price process.

In figure 4.1 we have simulated a few paths using this method. It is interesting to note that some paths appear to almost exhibit jumps even though we know the paths theoretically should be continuous. What happens is that the roughness property of volatility essentially allows for high spikes in volatility to occur over a short span of time thus allowing the price process to exhibit something that almost looks like a jump. In the next chapter we will see how this behaviour lets the model produce some empirically relevant implied volatility smiles.

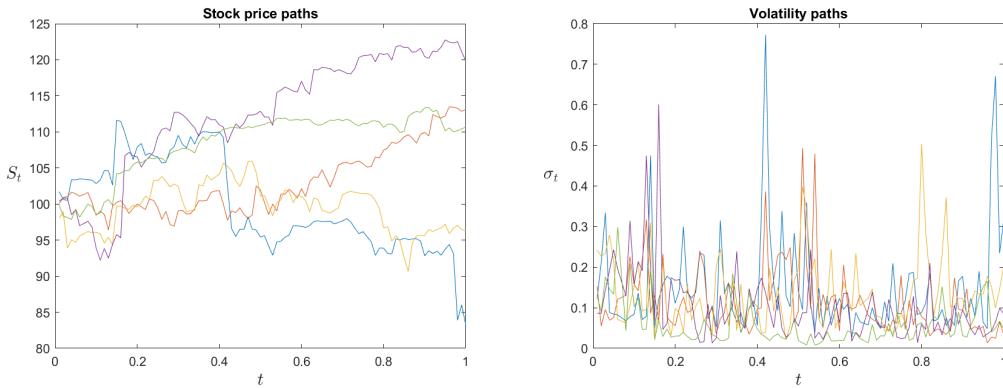


Figure 4.1: Sample paths from the rough Bergomi model with the parameters $H = 0.1, \rho = -0.9, \eta = 1.9, S_0 = 100, \xi_0(t) = 0.04$ for all t and $r = q = 0$.

Also, simulating 100,000 paths across 500 steps we get an average runtime of 3.15 seconds (average of 100 repetitions).⁴ While this seems reasonable even if we want to calibrate the model we think we can do better. This is why we consider a hybrid scheme in the next section.

4.3 Hybrid Scheme

In order to improve the performance of the simulation scheme without compromising too much on the distributional properties we will now follow (Bennedsen et al. 2017b) and present a hybrid scheme for simulation of the rough Bergomi model. While the paper derives a scheme for a more general class of processes, namely both Brownian semistationary (BSS) and truncated BSS processes, we

³This can be accomplished as follows: First find a lower diagonal matrix A s.t. $AA^\top = \Sigma$. This can be achieved using Cholesky factorisation. Next simulate a vector of $2M$ i.i.d. standard normals, call it Z . Finally return AZ . Standard arguments proves AZ will have the correct distribution.

⁴All simulation schemes and price estimators we show runtimes for are implemented in Matlab and are run on the GPU. We use a laptop with NVIDIA's GeForce MX150 graphics card with 2 GB memory and that on Windows version 10.0. For comparison we tried using our 1.6 - 3.4 GHz Intel Core i5 8250U CPU with 8 GB RAM but generally found this to be about 5-10 times slower.

will only consider the special case of simulating the rough Bergomi model which boils down to the question of simulating a special case of a truncated BSS process.

Let us first decompose \tilde{W}_0 at each of the time points as

$$\tilde{W}_0(t_i) = \sqrt{2H} \sum_{k=1}^i \int_{\frac{i-k}{n}}^{\frac{i-k+1}{n}} \left(\frac{i}{n} - s \right)^{-\gamma} dW_{1,s} \quad (4.11)$$

for $i = 1, 2, \dots, [nT]$.

To add some perspective on what we have done so far: Using the naive Riemann scheme from section 4.1 we approximated each of the above sub-integrals by assuming the integrand constant across each interval. However, recalling more generally that $\tilde{W}_0(t) = \sqrt{2H} \int_0^t (t-s)^{-\gamma} dW_{1,s}$ we see that the integrand $(t-s)^{-\gamma}$ explodes for s close to t . The constant approximation will therefore be inaccurate across intervals where this is the case. While the exact scheme avoids this problem by simultaneously simulating the entire process of \tilde{W}_0 on the time points it is also potentially slower for this reason. Thus we will here explain a hybrid method where only the first couple of problematic sub-integrals in (4.11) are simulated exactly, the remaining ones being approximated with a constant integrand. Let therefore $\kappa \in \{1, 2, 3, \dots\}$ be the number of sub-integrals we simulate exactly and let us approximate the remaining sub-integrals by using the approximation $(t-s)^{-\gamma} \approx \left(\frac{b_k}{n}\right)^{-\gamma}$ for $t-s \in [\frac{k-1}{n}, \frac{k}{n}]$ where the points $b_k \in [k-1, k]$, $k = \kappa + 1, \dots, [nT]$, can be chosen freely in each such interval.

The following approximation is therefore proposed:

$$\tilde{W}_0(t_i) \approx \sqrt{2H} \left(\sum_{k=1}^{\min\{i, \kappa\}} \underbrace{\int_{\frac{i-k}{n}}^{\frac{i-k+1}{n}} \left(\frac{i}{n} - s \right)^{-\gamma} dW_{1,s}}_{=: W_{i-k,k}^n} + \sum_{k=\kappa+1}^i \left(\frac{b_k}{n} \right)^{-\gamma} \underbrace{\int_{\frac{i-k}{n}}^{\frac{i-k+1}{n}} dW_{1,s}}_{=: W_{i-k}^n} \right) \quad (4.12)$$

for $i = 1, 2, \dots, [nT]$. Here the second sum should be treated as a zero if $i < \kappa + 1$.

As we have some freedom to choose the evaluation points $(b_k)_{k=\kappa+1}^{[nT]}$ we can choose them so as to make the scheme most accurate in an appropriate sense. In fact, proposition 2.8 in (Bennedsen et al. 2017b) shows that letting

$$b_k = \left(\frac{k^{1-\gamma} - (k-1)^{1-\gamma}}{1-\gamma} \right)^{-1/\gamma} \quad (4.13)$$

will minimize the asymptotic mean squared error.

With the above scheme what needs to be simulated are the $\kappa + 1$ dimensional vectors

$$\mathbf{W}_i^n = (W_i^n, W_{i,1}^n, \dots, W_{i,\kappa}^n) \quad (4.14)$$

for $i = 0, \dots, [nT] - 1$.

Luckily these are independent vectors as they deal with intervals of the stochastic integral that are non-overlapping and furthermore they are identically (and normally) distributed with a $(\kappa + 1) \times (\kappa + 1)$ dimensional covariance matrix Σ with entries given by

$$\Sigma_{1,1} = \frac{1}{n}. \quad (4.15)$$

and

$$\Sigma_{1,j} = \Sigma_{j,1} = \frac{(j-1)^{1-\gamma} - (j-2)^{1-\gamma}}{(1-\gamma)n^{1-\gamma}} \quad (4.16)$$

$$\Sigma_{j,j} = \frac{(j-1)^{1-2\gamma} - (j-2)^{1-2\gamma}}{(1-2\gamma)n^{1-2\gamma}} \quad (4.17)$$

for $j = 2, \dots, \kappa + 1$ and

$$\Sigma_{j,k} = \frac{n^{2\gamma-1}}{1-\gamma} \left((j-1)^{1-\gamma}(k-1)^{-\gamma} {}_2F_1 \left(1, \gamma, 2-\gamma, \frac{j-1}{k-1} \right) - (j-2)^{1-\gamma}(k-2)^{-\gamma} {}_2F_1 \left(1, \gamma, 2-\gamma, \frac{j-2}{k-2} \right) \right) \quad (4.18)$$

for $j, k = 2, \dots, \kappa + 1$, $j < k$, for the lower part of the remaining sub matrix.

A derivation is provided in the appendix.

To simulate $\tilde{W}_0(t)$ we may then independently simulate \mathbf{W}_i^n for each $i = 0, \dots, \lfloor nT \rfloor - 1$ and finally compute $\tilde{W}_0(t)$ via equation (4.12) and then the variance process via (4.3). Also, as \mathbf{W}_i^n is normally distributed you can use Cholesky factorization to simulate it. Having simulated $(\mathbf{W}_i^n)_{i=0}^{\lfloor nT \rfloor - 1}$ and using that

$$W_{1,\frac{i}{n}} = \sum_{j=0}^{i-1} W_j^n \quad (4.19)$$

for $i = 1, 2, \dots, \lfloor nT \rfloor$ we may then simulate the other independent Brownian motion $W_{\perp,t}$ on the same time points and construct the correlated Brownian motion as

$$W_{2,t} = \rho W_{1,t} + \sqrt{1-\rho^2} W_{\perp,t}. \quad (4.20)$$

The price process can then be constructed using a simple Euler scheme as in equation (4.2).

As is shown in (McCrickerd & Pakkanen 2018) the case $\kappa = 1$ is sufficiently accurate even for very rough processes and since we confirm this result in our own tests (see section A.6 in the appendix) and since higher values increases the computation time we also choose $\kappa = 1$ going forward. In this special case we get the following

$$\tilde{W}_0(t_i) \approx \sqrt{2H} \left(W_{i-1,1}^n + \sum_{k=2}^i \left(\frac{b_k}{n} \right)^{-\gamma} W_{i-k}^n \right) \quad (4.21)$$

for $i = 0, \dots, \lfloor nT \rfloor - 1$ and the independent vectors to simulate are only 2-dimensional.

Let us briefly compare the running times between the exact scheme and the hybrid scheme. The exact scheme: If we here use Cholesky factorisation to simulate the vector in (4.5) we see that this requires the multiplication of a lower diagonal $2n \times 2n$ matrix with a $2n$ dimensional vector. The total number of computations is then of order $\mathcal{O}(n^2)$. Technically Cholesky factorisation of Σ costs $\mathcal{O}(n^3)$ but it only needs to be evaluated once for all paths and so is of less importance in practise. Thus since the remaining computations are of order $\mathcal{O}(n)$ we should expect the total running time to be of order $\mathcal{O}(n^2)$. The hybrid scheme: For a general κ we may note that the second sum in (4.12) can be seen as a discrete convolution between two vectors. Efficient algorithms, such as fast Fourier transform (FFT), exists for evaluating such convolutions in $\mathcal{O}(n \log n)$ time. A straightforward computation of the sum for all relevant i would instead take $\mathcal{O}(n^2)$ time. Again, as the remaining computations are of order $\mathcal{O}(n)$ we should expect total running time to scale as $\mathcal{O}(n \log n)$. In conclusion we should expect a properly implemented hybrid scheme to be faster than

using exact simulation. However, using Matlab's own FFT routines and $n = 500$ we don't find a FFT implementation of (4.12) to be faster than using Matlab's **conv2** function. Thus we use the latter in practise. With this implementation simulating 100.000 paths with 500 steps using the Hybrid scheme with $\kappa = 1$ takes an average of 1.59 seconds (average of 100 repetitions) thus beating the exact scheme by roughly a factor 2. This is not a speed up orders of magnitude larger but still enough to make the implementation worthwhile. It is conceivable that the benefits of FFT will be more apparent if you simulate with more steps per year but then again we don't think one would practically need much more than what we used here.

4.4 Variance Reduction

In this section we consider a couple variance reduction methods as presented in (McCrickerd & Pakkanen 2018). The end goal is to estimate an implied volatility surface. Thus we will focus on call and put options. However, for simplicity of the exposition we assume we are standing at time zero and that we consider specifically a call option. The methods we propose are easily modified for the case of a put option.

As a start note that an application of Ito's lemma on $f(S_t) = \log(S_t)$ reveal the representation

$$S_t = S_0 \cdot \mathcal{E} \left(\int_0^t \sqrt{v_s} dW_{2,s} \right) \quad (4.22)$$

for $t \geq 0$. The implication is that the price of a call option with expiry $T > 0$ and strike K can be written as

$$E^Q \left((S_T - K)^+ \right) = S_0 \cdot E^Q \left(\left(\mathcal{E} \left(\int_0^T \sqrt{v_s} dW_{2,s} \right) - e^k \right)^+ \right) \quad (4.23)$$

with $k := \log(K/S_0)$ being log-moneyness.⁵ We may thus without loss of generality, and in addition to assuming zero interest rate and dividend, also assume $S_0 = 1$.

To be clear: Going forward we assume

$$S_t = \mathcal{E} \left(\int_0^t \sqrt{v_s} dW_{2,s} \right). \quad (4.24)$$

for $t \geq 0$. In terms of notation we will also write the price of a call option with expiry $T > 0$ and log-moneyness k as

$$P(k, T) := E^Q \left((S_T - e^k)^+ \right) \quad (4.25)$$

where S_T is then given by (4.24) with $t = T$.

A final comment before presenting our estimator: It is well-known that implied volatilities estimated by Monte Carlo simulation are significantly noisier when the underlying option is in-the-money rather than out-of-the-money. Thus for k negative we should instead use a put option and thus estimate $E^Q((e^k - S_T)^+)$ and compute the implied volatility based on that. This is what we have done in our own implementation.

To estimate $P(k, T)$ we limit ourselves to estimators of the form

$$\hat{P}_N(k, T) := \frac{1}{N} \sum_{i=1}^N (X_i + \hat{\alpha}_N Y_i) - \hat{\alpha}_N E(Y) \quad (4.26)$$

⁵In the general setting with non zero interest rate and dividend we define $k := \log(K/F_0)$.

where $(X_i, Y_i)_{i=1}^N$ are $N \in \mathbb{N}$ i.i.d. sampled random variables and $\hat{\alpha}_N$ the empirical version of $\frac{\text{Cov}(X, Y)}{\text{Var}(Y)}$ based on those N samples. The choice of $\hat{\alpha}_N$ is known to asymptotically minimize the variance of the estimator.⁶ Note that we write (X, Y) to denote a generic sample from $(X_i, Y_i)_{i=1}^N$.

Letting $X = (S_T - e^k)^+$ and $Y = 0$ we get the usual Monte Carlo estimator which we will also refer to as the *base* estimator. For the purpose of reducing variance we will however consider a specific *mixed* estimator that is a combination of using conditional Monte Carlo and control variates at the same time.⁷

Before presenting the mixed estimator we will decompose the price process $(S_t)_{t \geq 0}$ as follows:

$$S_t = \mathcal{E} \left(\int_0^t \sqrt{v_s} dW_{2,s} \right) \quad (4.27)$$

$$= \mathcal{E} \left(\rho \int_0^t \sqrt{v_s} dW_{1,s} + \sqrt{1 - \rho^2} \int_0^t \sqrt{v_s} dW_{\perp,s} \right) \quad (4.28)$$

$$= \mathcal{E} \left(\rho \int_0^t \sqrt{v_s} dW_{1,s} \right) \mathcal{E} \left(\sqrt{1 - \rho^2} \int_0^t \sqrt{v_s} dW_{\perp,s} \right) \quad (4.29)$$

$$=: S_t^1 \cdot S_t^\perp. \quad (4.30)$$

Here we have used the general result⁸

$$\mathcal{E}(Z^1) \cdot \mathcal{E}(Z^2) = \mathcal{E}(Z^1 + Z^2 + [Z^1, Z^2]) \quad (4.31)$$

for any continuous semi-martingales Z^1 and Z^2 , $[Z^1, Z^2]$ being the quadratic covariation process, and that for this particular case $W_{1,s} \perp\!\!\!\perp W_{\perp,s}$ such that the quadratic covariation between the two stochastic integrals is zero.

Let now $(\mathcal{F}_t^1)_{t \geq 0}$, respectively $(\mathcal{F}_t^\perp)_{t \geq 0}$, be the filtration generated by $(W_{1,t})_{t \geq 0}$, respectively $(W_{\perp,t})_{t \geq 0}$. Conditional on \mathcal{F}_T^1 the entire path $(v_t)_{0 \leq t \leq T}$ as well as S_T^1 will then be known.

⁶See (Glasserman 2004) section 4.1.

⁷In (McCrickerd & Pakkanen 2018) two other estimators, a conditional Monte Carlo without control variates and a control variate method without conditional Monte Carlo, are also proposed. However as the authors claim the *mixed* estimator is the most efficient one and so we only consider that estimator.

⁸Proof:

$$\begin{aligned} \mathcal{E}(Z_t^1) \cdot \mathcal{E}(Z_t^2) &= e^{Z_t^1 - Z_0^1 - \frac{1}{2}[Z^1]_t + Z_t^2 - Z_0^2 - \frac{1}{2}[Z^2]_t} \\ &= e^{(Z_t^1 + Z_t^2) - (Z_0^1 + Z_0^2) - \frac{1}{2}([Z^1]_t + [Z^2]_t)} \\ &= e^{(Z_t^1 + Z_t^2 + [Z^1, Z^2]_t) - (Z_0^1 + Z_0^2) - \frac{1}{2}([Z^1 + Z^2 + [Z^1, Z^2]]_t)} \\ &= \mathcal{E}(Z_t^1 + Z_t^2 + [Z^1, Z^2]_t) \end{aligned}$$

where the second last line follows since

$$[Z^1]_t + [Z^2]_t = [Z^1 + Z^2 + [Z^1, Z^2]]_t - 2[Z^1, Z^2]_t.$$

which follows by the computation

$$[Z^1 + Z^2 + [Z^1, Z^2]]_t = [Z^1 + Z^2]_t = [Z^1]_t + [Z^2]_t + 2[Z^1, Z^2]_t.$$

Since we can write

$$\log S_T^\perp = \log \mathcal{E} \left(\sqrt{1 - \rho^2} \int_0^T \sqrt{v_t} dW_{\perp,t} \right) \quad (4.32)$$

$$= \sqrt{1 - \rho^2} \int_0^T \sqrt{v_t} dW_{\perp,t} - \frac{1}{2} E \left(\left(\sqrt{1 - \rho^2} \int_0^T \sqrt{v_t} W_{\perp,t} \right)^2 \right) \quad (4.33)$$

$$= \sqrt{1 - \rho^2} \int_0^T \sqrt{v_t} dW_{\perp,t} - \frac{1}{2}(1 - \rho^2) \int_0^T v_t dt \quad (4.34)$$

we may conditionally on \mathcal{F}_T^1 conclude

$$\log S_T = \log S_T^1 + \log S_T^\perp \sim \mathcal{N} \left(\log S_T^1 - \frac{1}{2}(1 - \rho^2) \int_0^T v_t dt, (1 - \rho^2) \int_0^T v_t dt \right). \quad (4.35)$$

Let further $\text{BS}(\nu; s, k)$ be the Black-Scholes formula when inputting total variance ν , spot s , log strike k and keeping interest rates and dividends at zero.⁹

For the mixed estimator we define the random variable X as

$$X := E^Q \left((S_T - e^k)^+ | \mathcal{F}_T^1 \right) = \text{BS} \left((1 - \rho^2) \int_0^T v_t dt; S_T^1; k \right) \quad (4.36)$$

where the second equality follows since by (4.35) S_t is conditionally log-normal as under Black-Scholes. Furthermore an application of the Tower property for conditional expectations ensures $E^Q(X) = E^Q \left((S_T - e^k)^+ \right)$. This concludes the conditional Monte Carlo part of the estimator.¹⁰

Let now $V > 0$ be some constant. Since $dS_t^1 = \rho \sqrt{v_t} S_t^1 dW_{1,t}$ we have that $\Theta_t := \rho^2 \int_0^t v_s ds$ is the realized variance of S_t^1 . Defining further $V^* := \rho^2 V$ we let the control variate Y be given as

$$Y := E^Q \left((S_{\tau_{V^*}}^1 - e^k)^+ | \mathcal{F}_T^1 \right) \quad (4.37)$$

where $\tau_{V^*} = \inf \{t > 0 : \Theta_t = V^*\}$.

Assuming $\Theta_T < V^*$ one may note that Y is actually the time T price of a so-called *timer option* on the process $(S_t^1)_{t \geq 0}$ and that with a variance budget of $V^* - \Theta_T$ in excess of the already realized variance Θ_T . Since we assumed zero interest rate and dividend yield, and keeping the assumption $\Theta_T < V^*$, we can express this price as $\text{BS}(V^* - \Theta_T; S_T^1, k)$. See appendix B.3 for a proof of this result.

More generally, taking appropriate care of the possibility of $V^* \leq \Theta_T$, we get

$$Y = \text{BS}(V^* - \Theta_{T \wedge \tau_{V^*}}; S_{T \wedge \tau_{V^*}}^1, k). \quad (4.38)$$

Finally, as Y is a conditional expectation we find its expectation by using the Tower property for conditional expectations and combine that with (4.38) and (4.379 where we let $T = 0$):

$$E^Q(Y) = E^Q \left((S_{\tau_{V^*}}^1 - e^k)^+ \right) = \text{BS}(V^*; S_0^1, k) = \text{BS}(V^*; 1, k) \quad (4.39)$$

⁹By *total variance* we mean the product of the Black-Scholes squared volatility and time-to-expiry.

¹⁰We expect this to reduce the variance due to Jensen's inequality for conditional expectations as $x \mapsto x^2$ is convex.

This almost completes the description of the mixed estimator.

Nonetheless, we have yet to decide on the choice of V . One thing to note in this context is that evaluation of (4.38) is potentially quite slow as one has to compute the stopping times τ_{V^*} path by path. As there is no upper bound on Θ_T there is no constant $V > 0$ that will ensure $V^* > \Theta_T$ always holds such that we can avoid evaluation of τ_{V^*} in determining Y .

However as is proposed in (McCrickerd & Pakkanen 2018) one can try using

$$V = \sup \left\{ \left(\int_0^T v_t dt \right)_{i=1}^N \right\} \quad (4.40)$$

which can only be established post simulation. Importantly it will result in $Y = \text{BS}(V^* - \Theta_T; S_T^1, k)$ which is faster to evaluate. The choice unfortunately implies that we no longer have any guarantee that $\hat{P}(k, T)$ is a consistent estimator of $P(k, T)$. Nonetheless, we tested the above choice against the base estimator which is unbiased and found no observable bias - see section A.7 in the appendix for the results. In our own implementation we therefore use (4.40).

At this stage one may furthermore note that the mixed estimator actually only depends on $W_{1,t}$ such that we save extra computational time not having to simulate the other Brownian motion also. In relation to this we use an extra variance reduction method by constructing antithetic sample paths by flipping the sign of $W_{1,t}$. This may not only reduce the overall variance of the price estimator due to negative correlation between the antithetic samples but it also halves the required number of paths to simulate although the computational effort also increases some.

Finally, while both the conditional Monte Carlo and the control variates method are guaranteed to reduce the variance of the estimator we refer the reader to (McCrickerd & Pakkanen 2018) for actual numerical test showing how well this particular specification of the estimator works for the rough Bergomi model. We furthermore remark that even with the variance reduction techniques added the computational time does not change much compared to simply simulating the stock using the hybrid scheme. Specifically we are able to produce an implied volatility surface $(k, T) \in \{-0.20, -0.19, \dots, 0.19, 0.20\} \times \{0.1, 0.2, \dots, 0.9, 1\}$ using 100.000 paths (50.000 of which are antithetic) and 500 steps per year in about 1.44 seconds. The main reasons why we are able to keep computational time down is 1) only having to simulate S^1 and not also S^\perp and 2) by being able to compute half of the paths using antithetic samples. With these runtimes in mind calibration should be feasible.

Chapter 5

Calibration

In this chapter we consider the problem of calibrating the model to market data. The parameters to calibrate are the three parameters H , ρ , η and finally the entire forward variance curve $t \mapsto \xi_0(t)$. The market data we consider are bid and ask prices on call and put options on SPX for various strikes and expiries.

The chapter is structured as follows: First we investigate how to estimate the forward variance curve using variance swaps and log-contracts. Then we consider the calibration of H , ρ and η . While the hybrid scheme and variance reduction techniques from the last chapter has reduced the computational time significantly the general optimization problem is three-dimensional and thus still time consuming to solve. We therefore propose a more efficient calibration method using various expansions as derived in (Euch et al. 2018) and (Bergomi & Guyon 2011). Finally we test the calibration method on actual market data.

5.1 Estimating The Forward Variance Curve

For the purpose of estimating the forward variance curve it is useful to first consider the variance swap contract which is a swap contract on the total realized variance across some period. That is, a maturity $T > 0$ variance swap first contracted at time zero pays, at time T , the difference between $\int_0^T \sigma_t^2 dt$ and some strike where the strike is chosen such that the contract is worth zero when first contracted. Specifically the strike of the variance swap for a new contract is

$$E^Q \left(\int_0^T \sigma_t^2 dt \right). \quad (5.1)$$

Let us already now define the expiry T implied variance swap volatility observed at some time $t < T$ as

$$\hat{\sigma}_T^{VS}(t) := \sqrt{\frac{1}{T-t} E^Q \left(\int_t^T \sigma_s^2 ds \middle| \mathcal{F}_t \right)} \quad (5.2)$$

since we will refer to this quantity many times later on. Note also that if we consider the variance swap curve, i.e. (5.1), as a function of T and then take the derivative we get the forward variance curve:

$$\frac{d}{dT} E^Q \left(\int_0^T \sigma_t^2 dt \right) = \frac{d}{dT} \int_0^T E^Q (\sigma_t^2) dt = E^Q (\sigma_T^2) = \xi_0(T). \quad (5.3)$$

Variance swaps are somewhat actively traded in practise and thus one can technically extract the forward variance curve from observations of traded variance swaps. However, we do not have access to such data and neither are such contracts as liquid as standard call and put options.

The main difficulty is thus how to price the variance swap when we only have knowledge of call and put options. For this purpose we will first rewrite the variance swap in terms of a so-called log-contract which is a simple claim with payoff

$$g(S_T) = \log\left(\frac{S_T}{F_0}\right) = \log\left(\frac{F_T}{F_0}\right) \quad (5.4)$$

where $F_t := F(t, T) = e^{\int_t^T (r(s) - q(s)) ds} S_t$ is the forward price at time t with expiry at time T .

5.1.1 Replicating The Variance Swap

Here we will rewrite the variance swap in terms of the log-contract and other terms which will reveal a replicating strategy. This is the content of the below theorem:

Theorem 5.1.1. *The variance swap can be replicated by static positions in a zero-coupon bond, the underlying asset, the log-contract and finally a continuously rebalanced position in a futures contract.¹ In symbols:*

$$\int_0^T \sigma_t^2 dt = 2\left(\frac{S_T}{F_0} - \log\left(\frac{F_T}{F_0}\right) - 1\right) - 2 \int_0^T \left(\frac{1}{F_0} - \frac{1}{F_t}\right) dF_t. \quad (5.5)$$

Proof. Consider the following transformation f

$$f(t, x) = f(x) = \frac{2}{T} \left(\frac{x}{F_0} - \log\left(\frac{x}{F_0}\right) - 1 \right) \quad (5.6)$$

and note that $\frac{\partial f}{\partial t} = 0$, $\frac{\partial f}{\partial x} = \frac{2}{T} \left(\frac{1}{F_0} - \frac{1}{x} \right)$, $\frac{\partial^2 f}{\partial x^2} = \frac{2}{T} x^{-2}$.

An application of Ito on $f(F_t)$ then gives us

$$df(F_t) = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dF_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} (dF_t)^2 \quad (5.7)$$

$$= \frac{2}{T} \left(\frac{1}{F_0} - \frac{1}{F_t} \right) dF_t + \frac{1}{2} \cdot \frac{2}{T} F_t^{-2} F_t^2 \sigma_t^2 dt \quad (5.8)$$

$$= \frac{2}{T} \left(\frac{1}{F_0} - \frac{1}{F_t} \right) dF_t + \frac{1}{T} \sigma_t^2 dt. \quad (5.9)$$

On integral form this reads

$$f(F_T) = f(F_0) + \frac{1}{T} \int_0^T \sigma_t^2 dt + \frac{2}{T} \int_0^T \left(\frac{1}{F_0} - \frac{1}{F_t} \right) dF_t \quad (5.10)$$

and since $f(F_0) = 0$ we can rearrange this to get

$$\int_0^T \sigma_t^2 dt = 2\left(\frac{S_T}{F_0} - \log\left(\frac{F_T}{F_0}\right) - 1\right) - 2 \int_0^T \left(\frac{1}{F_0} - \frac{1}{F_t}\right) dF_t. \quad (5.11)$$

□

¹As both the interest rate as well as the dividend yield on the underlying are deterministic it follows that the forward and futures prices are equal at all times. See (Lioui & Poncet 2005) chapter 2 for a proof. It is then easily shown that the last term in equation (5.5) is the gain from some continuously rebalanced trading strategy in the futures contract.

Taking Q expectation in (5.5) we see that the variance swap is exactly proportional to the price of the log-contract:

$$E^Q \left(\int_0^T \sigma_t^2 dt \right) = 2 \left(\frac{F_0}{F_0} - E^Q \left(\log \left(\frac{F_T}{F_0} \right) \right) - 1 \right) - 2 \cdot 0 \quad (5.12)$$

$$= -2E^Q \left(\log \left(\frac{F_T}{F_0} \right) \right). \quad (5.13)$$

Thus we can derive the variance swap curve if we can properly price the corresponding log-contracts.

5.1.2 Pricing The Log-Contract

Here we consider how to replicate and thus price the log-contract. First we will however prove a more general replication result. We state the result below in purely mathematical terms:

Theorem 5.1.2. *Assume $f : (0, \infty) \mapsto \mathbb{R}$ is some twice continuously differentiable function and let $x_0 > 0$. The following decomposition then holds:*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \int_0^{x_0} f''(z)(z - x)^+ dz + \int_{x_0}^\infty f''(z)(x - z)^+ dz. \quad (5.14)$$

Proof. Consider first the case of $x \geq x_0$. Then integration by parts gives us

$$\int_{x_0}^\infty f''(z)(x - z)^+ dz = \int_{x_0}^x f''(z)(x - z) dz \quad (5.15)$$

$$= [f'(z)(x - z)]_{x_0}^x - \int_{x_0}^x f'(z) \cdot (-1) dz \quad (5.16)$$

$$= -f'(x_0)(x - x_0) + f(x) - f(x_0). \quad (5.17)$$

The above integral is of course zero if $x < x_0$ though in this case we get

$$\int_0^{x_0} f''(z)(z - x)^+ dz = \int_x^{x_0} f''(z)(z - x) dz \quad (5.18)$$

$$= [f'(z)(z - x)]_x^{x_0} - \int_x^{x_0} f'(z) \cdot 1 dz \quad (5.19)$$

$$= -f'(x_0)(x - x_0) + f(x) - f(x_0). \quad (5.20)$$

In the case $x \geq x_0$ the above integral is zero.

In total we conclude

$$\int_0^{x_0} f''(z)(z - x)^+ dz + \int_{x_0}^\infty f''(z)(x - z)^+ dz = -f'(x_0)(x - x_0) + f(x) - f(x_0) \quad (5.21)$$

and rearranging we get the desired result. \square

Using theorem 5.1.2 with $f(x) = \log \left(\frac{x}{F_0} \right)$ and $x_0 = F_0$, noting that $f'(x) = x^{-1}$ and $f''(x) = -x^{-2}$, we get

$$f(S_T) = f(F_0) + f'(F_0)(S_T - F_0) - \int_0^{F_0} \frac{1}{K^2} (K - S_T)^+ dK - \int_{F_0}^\infty \frac{1}{K^2} (S_T - K)^+ dK \quad (5.22)$$

that is

$$\log\left(\frac{S_T}{F_0}\right) = \frac{S_T}{F_0} - 1 - \int_0^{F_0} \frac{1}{K^2} (K - S_T)^+ dK - \int_{F_0}^{\infty} \frac{1}{K^2} (S_T - K)^+ dK.$$

We see that the log-contract can be statically replicated by holding a position in the underlying stock, a zero-coupon bond and an infinite amount of put and call options of the same expiry each with weights K^{-2} .

Taking Q-expectation on both sides we get

$$E^Q\left(\log\left(\frac{S_T}{F_0}\right)\right) = -e^{\int_0^T (r(s)-q(s))dt} \left(\int_0^{F_0} \frac{1}{K^2} P(K, T) dK + \int_{F_0}^{\infty} \frac{1}{K^2} C(K, T) dK \right) \quad (5.23)$$

where $P(K, T)$, respectively $C(K, T)$, is the time zero price of an expiry T put, respectively call, option with strike K .

Summing up we have

$$E^Q\left(\int_0^T \sigma_t^2 dt\right) = 2e^{\int_0^T (r(s)-q(s))dt} \left(\int_0^{F_0} \frac{1}{K^2} P(K, T) dK + \int_{F_0}^{\infty} \frac{1}{K^2} C(K, T) dK \right). \quad (5.24)$$

The result above of course requires a full surface of call and put options which is at odds with the finitely many market observations that we have available. To mend this problem we will have to make an initial calibration to the observed market prices in order to interpolate and extrapolate a full surface of option prices. We discuss this later in section 5.3.

5.2 Market Data and Implied Yields

The option prices we consider in this thesis are bid and ask prices on SPX options from the 15th of September 2011.⁴ In total we have 1711 bid and ask prices available and that across 13 different expiries in the interval $[0.02, 2.27]$ as well as the index closing price for that date. We immediately remove two expiries in the data set as we find that expiries that are close together and have a moderate bid-ask spread relative to the time distance leads to an unreasonable forward variance curve between those expiries.² Let for future use \mathcal{T} be the set of the 11 remaining expiries. Before starting any analysis we furthermore remove options with bid prices equal to zero as they naturally contain no information value. Also, we decide to simply calibrate the model to the mid-prices, i.e. the average of the bid and ask prices. Unless otherwise specified we therefore refer to the mid-price when we talk about to an option price.

For simplicity we assume that the dividend yield $q(t) = q$ is constant whereas the interest rate may be time-dependent. Under this assumption the discount factor from equation (3.41) becomes $D(0, T) = e^{-(y(T)-q)T}$ where $y(T)$ is the maturity T yield-to-maturity. Thus, before we can calibrate the model we need to find the yield curve and dividend yield. We do so by finding all pairs of call and put options with the same strike and expiry. We find 619 such pairs and that roughly evenly distributed across the expiries. For these pairs the put-call parity must hold:

$$C(T, K) - P(T, K) = e^{-qT} S_0 + e^{-y(T)T} K. \quad (5.25)$$

²We are certain that this problem can be alleviated by properly handling the size of the bid-ask spread. However, as we think solving that problem is not central to the topic of this thesis we opt simply to remove the data.

The only unknown quantities in the above relation are q and $(y(T))_{T \in \mathcal{T}}$. Thus considering the relation for all strikes and expiries we wish to determine the dividend yield and zero coupon bond yields so as to minimize the distance between the left hand side and right hand side for all such put-call relations. In practise we minimize the sum of squared distances

$$\min_{q, (y(0, T))_{T \in \mathcal{T}}} \sum_{(T, K)} \left(C(T, K) - P(T, K) - e^{-qT} S_0 - e^{-y(T)T} K \right)^2 \quad (5.26)$$

where the sum is over all such observed put-call pairs.

Solving the above problem we get an implied dividend yield of $q = 0.0239$ which is roughly in line with historical values. The yield curve can be seen in figure 5.1.³ We think the yield curve is realistic enough as it is upwards sloping. Also, negative yields for short maturities are not unrealistic when compared to typical yield curves observed after the 2008 financial crisis.

In the remaining analysis we furthermore restrict ourselves to quotes coming from out-of-the-money options. We make this choice since out-of-the-money options typically are more liquid. Also, since we wish to calibrate the rough Bergomi model to the implied volatilities we have to decide whether or not to extract the observed implied volatility from the put or call option when we have both available. This leaves us with a total of 797 options going forward.

5.3 SVI Parameterization

In order to interpolate and extrapolate the observed option prices so as to calculate the initial forward variance curve we make use of the so-called stochastic volatility inspired (SVI) parametrization which is a parametric approach to modelling the implied volatility surface. The theory below and the implementation⁴ we use is mainly the result of the work presented in (Gatheral & Jacquier 2013).

Let $\sigma_{BS}(k, T)$ be the Black-Scholes implied volatility of an expiry T , log-moneyness k European option. The standard way of expressing a SVI parametrization is in terms of the total implied variance defined as $w(k, T) := \sigma_{BS}^2(k, T)T$. For fixed $T > 0$ we will refer to the mapping $k \mapsto \sigma_{BS}(k, T)$ as a single slice of the implied volatility surface. Equivalently we will refer to a slice of the total implied variance surface as the mapping $k \mapsto w(k, T)$ for a fixed $T > 0$.

³See section A.8 in the appendix for more details.

⁴We are thankful to Philipp Rindler from ETH Zürich to have made Matlab functionality that implements the SVI publicly available as well as having provided the data set of SPX options that we have used in this thesis. His code can be found at: https://se.mathworks.com/matlabcentral/fileexchange/49962-gatherals-and-jacquier-s-arbitrage-free-svi-volatility-surfaces?s_cid=ME_prod_FX

We will now consider the natural SVI parametrization which models each slice with a different set of parameters. For this reason we will drop the explicit reference to the expiry T and simply write $w(k; \mathcal{X})$ for a fixed total implied variance slice, where now \mathcal{X} is the set of parameters that models this particular slice. Under the SVI parametrization we model each slice through a parameter set $\mathcal{X}_N = (\Delta, \mu, \rho, \omega, \zeta)$ by assuming that the total implied variance slice has the explicit form

$$w(k; \mathcal{X}_N) = \Delta + \frac{\omega}{2} \left(1 + \zeta \rho (k - \mu) + \sqrt{(\zeta(k - \mu) + \rho)^2 + (1 - \rho^2)} \right). \quad (5.27)$$

To ensure non-negativity we require $\Delta, \mu \in \mathbb{R}$, $\omega \geq 0$, $|\rho| < 1$, $\zeta > 0$.

It is important to stress that there is no guarantee that any SVI parametrization of the above form will be free of arbitrage. However, certain extra restrictions can be imposed on the combined set of parameters to ensure no static arbitrage.

As the above parametrization is five-dimensional in each slice we wish to find an initial guess in a more restricted parameter space which hopefully will improve the robustness and speed of the optimization. We consider the so-called surface SVI (SSVI) parametrization which for the expiry $T \in \mathcal{T}$ slice restricts the parameters to the form $\mathcal{X}_0 = (0, 0, \rho, \theta_T, \phi(\theta_T))$. Here θ_T is a parameter specific for each slice $T \in \mathcal{T}$ whereas ρ is a common parameter for all slices and ϕ is some function satisfying certain regularity properties.⁵

Importantly, with the SSVI parametrization we get $w(0; \mathcal{X}_0) = \theta_T T$. That is, the at-the-money total variance in the SSVI is exactly $\theta_T T$. A natural way of choosing $(\theta_T)_{T \in \mathcal{T}}$ is therefore for each slice to put θ_T equal to the observed at-the-money implied variance which we find by using linear interpolation between the two quotes nearest at-the-money.

For ϕ we assume the explicit form

$$\phi(\theta_T) = \frac{\eta}{\sqrt{\theta_T(1 + \theta_T)}} \quad (5.28)$$

for some common parameter $\eta \geq 0$.

The optimization problem is now reduced to a two-dimensional problem in the parameters (ρ, η) . Also, from remark 4.4. in (Gatheral & Jacquier 2013), the specific SSVI parametrization we have chosen is free of static arbitrage exactly if the extra constraint $\eta(1 + |\rho|) \leq 2$ is satisfied.

For the objective function we choose to minimize the sum of squared differences in total implied variance between the observations and those produced by the SSVI. Constrained optimization with Matlab's `fmincon` function yields $(\rho, \eta) = (-0.73, 0.96)$. A few of the smiles of the initial calibration can be seen in figure 5.2. The rest can be found in section A.9 in the appendix. Unsurprisingly the fit is not perfect for all expiries in this restricted parametrization.

For the full calibration of the SVI it is more useful to express the parametric form in a different but equivalent set of parameters. For this purpose we define first the raw parameterization $\mathcal{X}_R = \{a, b, \rho, m, \sigma\}$ as

$$w(k; \mathcal{X}_R) = a + b \left\{ \rho(k - m) + \sqrt{(k - m)^2 + \sigma^2} \right\} \quad (5.29)$$

with the requirements $a \in \mathbb{R}$, $b \geq 0$, $|\rho| < 1$, $m \in \mathbb{R}$, $\sigma > 0$ to ensure positivity.

Inversion equations between the natural and raw parametrization can be found in (Gatheral &

⁵Again, we refer the reader to (Gatheral & Jacquier 2013) for the full details.

Jacquier 2013).

Define now also the SVI-Jump-Wings (SVI-JW) parametrization $\mathcal{X}_J = (v_T, \psi_T, p_T, c_T, \tilde{v}_T)$ for an expiry $T > 0$ slice as

$$v_T = \frac{a + b \{ -\rho m + \sqrt{m^2 + \sigma^2} \}}{T} \quad (5.30)$$

$$\psi_T = \frac{1}{\sqrt{w_T}} \frac{b}{2} \left(-\frac{m}{\sqrt{m^2 + \sigma^2}} + \rho \right) \quad (5.31)$$

$$p_T = \frac{1}{\sqrt{w_T}} b(1 - \rho) \quad (5.32)$$

$$c_T = \frac{1}{\sqrt{w_T}} b(1 + \rho) \quad (5.33)$$

$$\tilde{v}_T = \frac{a + b\sigma\sqrt{1 - \rho^2}}{T} \quad (5.34)$$

with $w_T := v_T T$ where we have added explicit dependence on the expiry T . An inversion formula in (Gatheral & Jacquier 2013) shows that this parametrization is completely equivalent to the raw parameterization and thus also the natural one. The reason for making the parameters expiry dependent is that they then contain a better economic intuition. For instance v_T is the at-the-money variance and ψ_T the at-the-money skew multiplied by \sqrt{T} . The first claim is easily seen by comparing (5.30) with (5.29) letting $k = 0$.

The last claim we prove here:

$$\frac{d\sigma_{BS}(k, T)}{dk} \Big|_{k=0} = \frac{d}{dk} \left(\sqrt{\frac{w(k, T)}{T}} \right) \Big|_{k=0} \quad (5.35)$$

$$= \frac{1}{2\sqrt{T}} (w(0, T))^{-\frac{1}{2}} b \left(\rho + \frac{1}{2} ((0 - m)^2 + \sigma^2)^{-\frac{1}{2}} \cdot 2 \cdot (0 - m) \right) \quad (5.36)$$

$$= \frac{1}{\sqrt{T}} \frac{1}{\sqrt{w(0, T)}} \frac{b}{2} \left(-\frac{m}{\sqrt{m^2 + \sigma^2}} + \rho \right) \quad (5.37)$$

$$= \frac{\psi_T}{\sqrt{T}}. \quad (5.38)$$

Under this parametrization no butterfly arbitrage can be ensured by restricting the last two parameters as

$$c_T = p_T + 2\psi_T \quad (5.39)$$

$$\tilde{v}_T = v_T \frac{4p_T c_T}{(p_T + c_T)^2}. \quad (5.40)$$

We stress that this requirement can be relaxed to something more general but we think the above restriction is easy to use and gives us decent results. Finally as is shown in (Gatheral & Jacquier 2013), to avoid calendar arbitrage it is enough to require $\frac{\partial w(k, t)}{\partial t} \geq 0$.

In light of the above we now propose a recipe for calibrating the full SVI which will then ensure no static arbitrage. Loop for all $T \in \mathcal{T}$ in descending order and do the following: Calibrate first the three parameters (v_T, ψ_T, p_T) by minimizing the square root of the sum of squared distances between the total implied variances observed and those produced by the SVI but with a heavy penalty for crossing above the previous total implied variance smile. As initial guesses to this optimization

we use the parameters obtained from the SSVI. Finally we set (c_T, \tilde{v}_T) using equations (5.39) - (5.40).

We are well aware that many other recipes can be proposed but we find the above to work reasonably well as can be seen from the fit in figure 5.2. The fitted parameters can be found in section A.9 in the appendix.

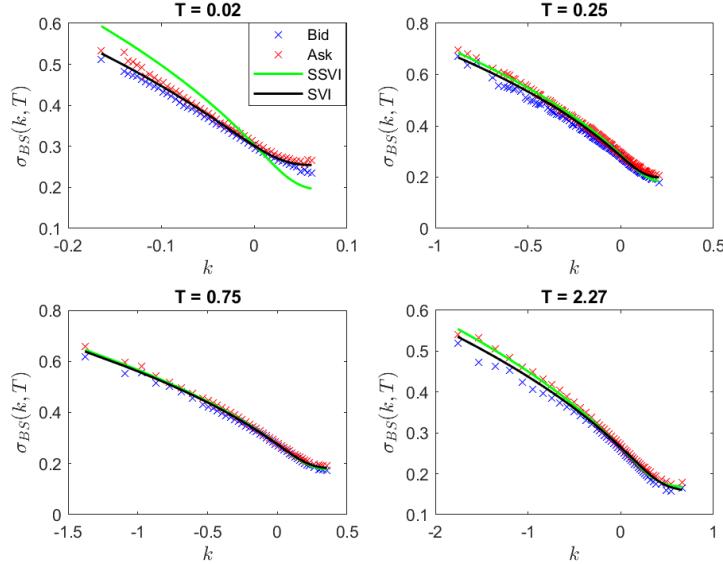


Figure 5.2: *SSVI and SVI fits to SPX options data from the 15th of September 2011. Plots for the remaining expiries can be found in section A.9 in the appendix*

With the SVI fit we now have a full smile available for each expiry in the data set. Using a Riemann sum approximation to (5.24) we may therefore calculate the variance swap curve and by differentiation thereafter the forward variance curve. These are shown in figure 5.3. For simplicity we have chosen to interpolate and extrapolate the forward variance curve flat between the observed dates. Other methods could be used but this choice makes it easy to readjust the forward variance curve when we later attempt to iterate on the curve in order to match the observed at-the-money implied volatilities within the rough Bergomi model.

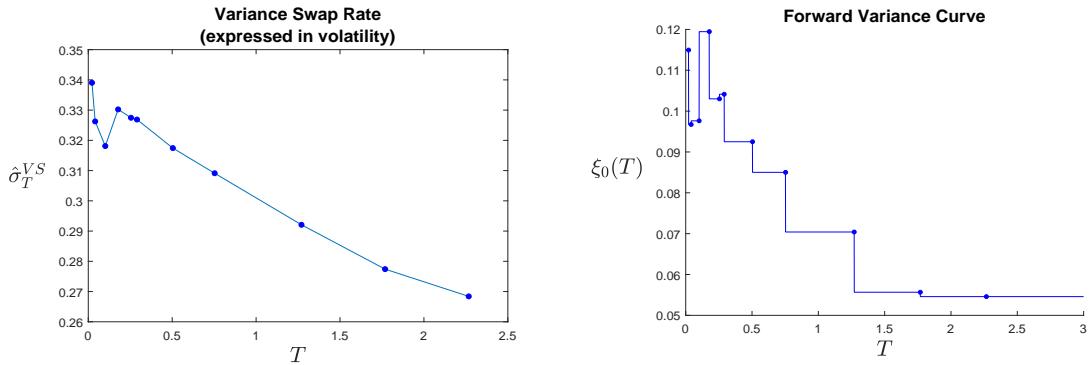


Figure 5.3: The variance swap rate expressed in volatility, i.e. $\hat{\sigma}^{VS}(T)$, and the forward variance curve $t \mapsto \xi_0(t)$ both as extracted from the SVI fit.

As a final application of the SVI fit let us briefly consider the term structure of the at-the-money skew that it proposes. Since $\psi(T) := \frac{d\sigma_{BS}(k,T)}{dk}|_{k=0} = \psi_T \cdot T^{-1/2}$ this information is easily extracted. We plot the absolute value of the at-the-money skew in figure 5.4 together with a power-law fit.

As we can see the skew increases rapidly in absolute value for short expiries and the data seems very well explained by the power-law fit. This is exactly the same thing documented in (Gatheral et al. 2017) and one of the big motivations behind finding alternative models to standard stochastic volatility models that are unable to produce this behaviour. While the rapidly increasing skew can be achieved by adding jumps another possibility is the rough Bergomi model with its rough volatility.⁶ We will see why when we calibrate the model in the next section.

5.4 Calibrating the rough Bergomi model

Various approaches can be taken to calibrate the model to the observed option prices. The first approach we considered is to minimize the sum of squared differences in implied volatilities, i.e. the expression⁷

$$\sum_{(k,T)} (\sigma_{BS}^{obs}(k,T) - \sigma_{BS}^{model}(k,T))^2, \quad (5.41)$$

with respect to the entire parameter vector (H, ρ, η) and that subject to the constraints $H \in (0, \frac{1}{2})$, $\rho \in [-1, 1]$ and $\eta > 0$. We use Matlabs **fminsearch** function with a heavy penalty for violating the parameter bounds. However, having completed this optimization step it is far from guaranteed that the fitted model will match the empirical at-the-money implied volatilities. To mend this problem we iterate on the forward variance curve, adjusting one of the flat sections at the time until all at-the-money implied volatilities are matched. Unfortunately even though we attempted to use several different starting values for the optimization the optimizer did not converge to any reasonable solution with a good fit. In fact it often stays very close to the initial value we provide it perhaps indicating that the function we are trying to optimize contains many local minima. With such a method we think one would have to supply a very large amount of initial guesses thus making the total computation very heavy. Our own experience is therefore that three-dimensional optimization is infeasible. To arrive at a faster and more robust method we now propose a fast calibration scheme that we have developed using various approximations found in the literature.

Firstly, from the Bergomi-Guyon implied volatility expansion formula presented in (Bergomi &

⁶While jump models may fit the term structure of skew well, we will however in chapter 6 argue that jump models are generally unable to properly capture the dynamic properties of the implied volatility surface. As we will also argue, the rough Bergomi model can also fit this empirical fact well.

⁷Here $\hat{\sigma}_{BS}^{obs}(k, T)$ is the observed implied volatility for a strike K , expiry $T > 0$ and $\hat{\sigma}_{BS}^{model}(k, T)$ is the corresponding implied volatility under the rough Bergomi model.

Guyon 2011) you can show that the at-the-money skew under the rough Bergomi model is approximately given by⁸

$$\psi(T) = \frac{\rho\tilde{\eta}}{2}[(1-\gamma)(1-\gamma)]^{-1}T^{-\gamma} \quad (5.42)$$

where $\tilde{\eta} := \sqrt{2H} \cdot \eta$. This is exactly a power-law relationship as proposed by the empirical observations in figure 5.4. Importantly the formula (5.42) suggests that this relationship is an inherent property of the model.

Taking absolute value and log we get the relationship

$$\log(|\psi(T)|) = C - \gamma \log(T) \quad (5.43)$$

where C is a specific, but in this context irrelevant, constant. Considering the above equation for all $T \in \mathcal{T}$ we can infer the value of γ via OLS regression. On the SPX options data doing so yields exactly the same coefficients as the fit in figure 5.4. Specifically $\gamma = 0.415$ or $H = \frac{1}{2} - \gamma = 0.085$. As this value corresponds well to the estimates of roughness obtained in chapter 2 there, at least initially, seems to be a correspondence between the dynamics of volatility under P and option prices under Q using this model.

Having fixed H , we are now left with a two-dimensional optimization problem in (ρ, η) . We will now attempt to further reduce this to a one dimensional problem by letting $\rho = \rho(\eta)$ be a function of η . Specifically, letting $\eta > 0$ be given we will choose ρ so as to always match the observed at-the-money skew $\psi(T^*)$ for some chosen expiry $T^* \in \mathcal{T}$. We propose this form of choosing ρ as we find that this is the main parameter controlling the at-the-money skew of the smile for a fixed expiry.⁹ Thus if we have a method of computing the at-the-money skew under the model we can simply use an optimization algorithm to iterate to the correct level of ρ . To compute the at-the-money skew one could use a finite difference approximation to the derivative. Unfortunately, this would necessarily have to be combined with Monte Carlo estimation of implied volatilities and may thus be quite slow. To alleviate this last problem we have instead considered various expansion formulas for the at-the-money skew.

We will therefore now review the three approximations that we have considered: One is the formula stated in (5.42). This is the first order approximation to at-the-money skew that the Bergomi-Guyon expansion suggests.

The Bergomi-Guyon expansion formula also contains a second order approximation to $\psi(T)$. For a general forward variance curve that approximation will have to be computed using numerical integration. However for a flat forward variance curve, i.e. assuming $\xi_0(t) = \xi_0(0)$ for all t , the formula suggests

$$\psi(T) = \hat{\sigma}_T^{VS} \left[\frac{1}{2v^2} C^{x\xi} + \frac{1}{8v^3} (4C^\mu v - 3(C^{x\xi})^2) \right] \quad (5.44)$$

where

$$C^{x\xi} = \eta^2 \xi_0(0)^2 D_H^2 \frac{T^{2+2H}}{2+2H} \quad (5.45)$$

and

$$C^\mu = \frac{1}{2} \rho^2 \eta^2 \xi_0(0)^2 D_H^2 \left(1 + \frac{\Gamma(H+3/2)^2}{\Gamma(2H+3)} \right) \frac{T^{2+2H}}{2+2H} \quad (5.46)$$

⁸We cite their main result in section B.6 in the appendix and derive the mentioned formula.

⁹We provide a few plots illustrating the impact of varying each of the three parameters in section A.10 in the appendix.

with $D_H := \frac{\sqrt{2H}}{H+\frac{1}{2}}$ as well as $v := \int_0^T \xi_0(t) dt = T\xi_0(0)$ s.t. also $\hat{\sigma}_T^{VS} = \sqrt{\frac{v}{T}} = \sqrt{\xi_0(0)}$.

The final skew approximation we consider is an expression that was proposed in (Fukasawa 2017) and is claimed to occur in one of the proofs of (Fukasawa 2015) where a short expiry expansion of implied volatilities in a general class of rough volatility models is considered. The approximation states

$$\psi(T) = -\frac{\rho}{\sqrt{T}} E \left(\frac{X_T}{\sqrt{\langle X \rangle_T}} \right) \quad (5.47)$$

where $\langle \cdot \rangle_T$ is the quadratic variation process and we have defined

$$X_T := \int_0^T \sqrt{v_t} dW_{1,t}. \quad (5.48)$$

The expression is unfortunately only semi-analytical as no explicit expression for the expectation in (5.47) is available and so naturally that part will have to be estimated using Monte Carlo.¹⁰

In figure 5.5 (see the next page) we compare the various approximative formulas with the actual at-the-money skew as computed using a finite difference derivative with Monte Carlo simulation. We see that while the second order Bergomi-Guyon skew expansion formula does better than that of first order the error is still quite noticeable across most expiries. On the other hand we see that the formula (5.47) proposed in (Fukasawa 2017) appears very accurate at all expiries although best for very small expiries. The last part is not completely unsurprising since the proofs in (Fukasawa 2015) considers approximations in the limit $T \rightarrow 0$.

Per the above we therefore settled on the expression (5.47). While the expression requires Monte Carlo simulation to be estimated it is extremely useful for our purposes for a particular reason. Note that what appears within the expectation does not depend on ρ . Thus having fixed η (and H) we can run a single Monte Carlo estimation of the expectation and immediately set¹¹

$$\rho(\eta) := -\sqrt{T^*} \cdot \psi(T^*) \cdot \left(\hat{E} \left(\frac{X_{T^*}}{\sqrt{\langle X \rangle_{T^*}}} \right) \right)^{-1} \quad (5.49)$$

which will then make sure that we match the at-the-money skew at the expiry T^* . For certain choices of parameters, expiry and observed skew it is possible that the above expression yields $\rho > 1$ or $\rho < -1$. In that case we choose to truncate the value to either $\rho = -1$ or $\rho = 1$ whichever is the closest.

¹⁰Compared to using a basic finite difference approximation to $\psi(T)$ there is still a notable improvement in speed using (5.47) as we avoid simulation of the stock price as well as the estimation of the implied volatilities altogether. Also as we will argue later the formula is semi-analytical in just the right way for our purpose.

¹¹Here \hat{E} indicates that the expectation is estimated using Monte Carlo.

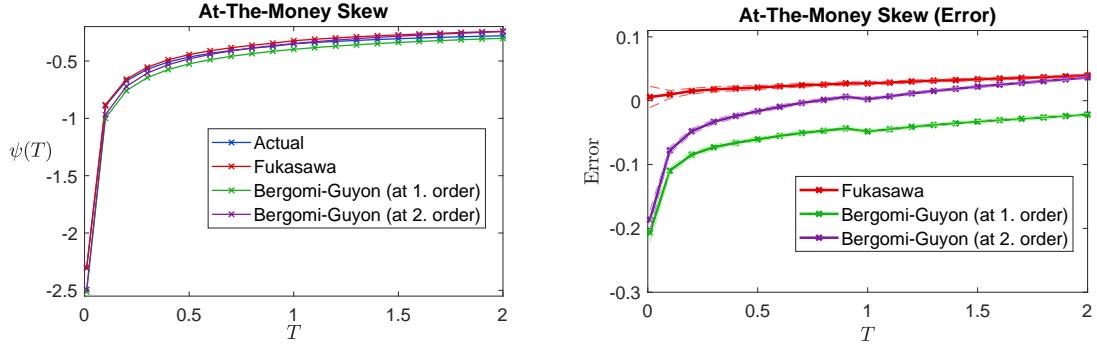


Figure 5.5: The term structure of at-the-money skew $\psi(T)$ as computed using various methods. The model parameters are: $H = 0.1$, $\eta = 1.9$, $\rho = -0.9$, $\xi_0(t) = 0.04$ for all t , $r = 0.05$, $q = 0.02$. The Monte Carlo average is based on a finite difference approximation to the derivative where the implied volatilities are computed using Monte Carlo. The shown curve is the average of 100 such repeated estimates each based on 10,000 paths. The Fukasawa formula is also computed using the average of 100 estimates each based on 10,000 sample paths. Dashed lines are 95 % confidence intervals computed using 100,000 bootstrapped samples based on the 100 estimates.

With the above scheme in place we are left with simply maximizing (5.41) with respect to η where H is already fixed and $\rho = \rho(\eta)$ as above. This is a one dimensional problem which we solve using Matlab's function **fminbnd**. We choose $T^* = 0.02$ as this is the shortest expiry in our data set and we know the skew approximation is best for short expiries. With this method we arrive at the final estimates $H = 0.0850$, $\rho = -0.9185$, $\eta = 1.9859$ and get a Root-Mean-Squared-Error (RMSE) of 0.0087.¹² We plot the fit in figure 5.6.

We see that the model fits the market data well across all expiries. In particular it is able to capture the power-law decay of skew which other models such as Heston is unable to do. Also the calibrated value of H is well in line with the values we estimated under the historical probability measure in chapter 2. This adds another layer of validity for the rough Bergomi model. We do however remark that being able to fit an implied volatility surface is not enough to have a good model. In practise the dynamics properties of the implied volatility surface is at least as important as the dynamic properties have large implications for hedging. This is a point we explore in chapter 6.

In general we conclude our scheme works quite well for obtaining a good fit to the data. We leave it to future research to document the general performance and practical applicability of our new calibration method and to see if one can implement a refined version that results in more realistic parameters without obtaining a worse fit. At the very least the algorithm can be used to quickly find an initial guess that can be fed to a more general optimization algorithm. Also for the data we considered the part of the algorithm that estimates (H, ρ, η) only took 126 seconds making it computationally much more feasible than solving the three-dimensional optimization problem.¹³

¹²We define the root-mean-squared error (RMSE) as $RMSE := \sqrt{\frac{1}{N_{obs}} \sum_{(k,T)} (\hat{\sigma}_{BS}^{obs}(k,T) - \hat{\sigma}_{BS}^{model}(k,T))^2}$ where N_{obs} is the total number of observed options we calibrate to.

¹³If we include fitting the SVI and iterating on the forward variance curve it took 539 seconds.

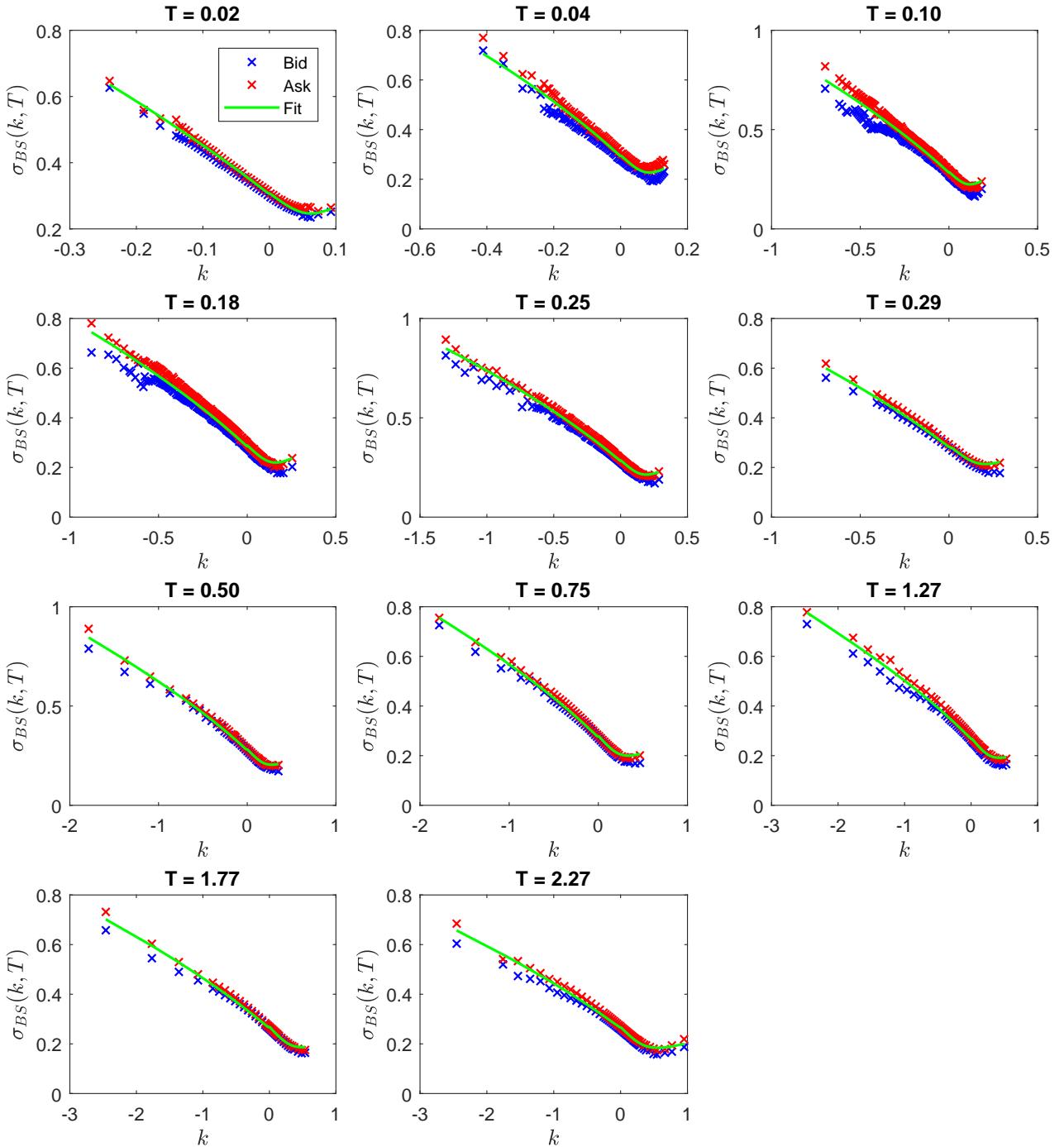


Figure 5.6: The rough Bergomi model fitted to SPX options data from the 15th of September 2011 using the fast calibration method. To estimate implied volatilities for both calibrations we used the mixed estimator with 100,000 paths (50,000 of which are antithetic).

Chapter 6

Hedging

In chapter 1 we saw how the Hurst exponent H varies moderately across time. In light of this empirical fact we find it reasonable to believe that option traders may have differing opinions on what value of H is the correct one to be used for options pricing. The main motivation behind this chapter is therefore to investigate and present thoughts on how one can realize a PnL in a situation where the Hurst exponent implied by options prices differs from the Hurst exponent that governs the actual underlying dynamics.

The chapter is structured as follows: We first look at hedging in the rough Bergomi model. This is non-trivial as the pricing function is not Markov in the stock price and spot volatility due to the non-Markovian nature of the fractional Brownian motion driving volatility. However, the pricing function is instead Markov in the stock price and the forward variance curve. This leads to a natural hedging strategy using the forward variance curve. With the initial goal of isolating the difference between the actual and market implied Hurst exponents as a PnL we next consider the so-called skew-stickiness-ratio (SSR) which is connected to the dynamic properties of the model. Under the rough Bergomi we find that this quantity is essentially equal to $\frac{3}{2} + H$. Since we find betting on H to be more naturally formulated as how one can bet on the difference between realized and implied SSR we lastly present some initial thoughts on how this can be done.

A technical note: In the following chapter we will often write $\xi_t := (\xi_t(u))_{u>t}$ to refer to the entire forward variance curve observed at time t .

6.1 Hedging Strategy

Before deriving the hedging strategy we need to rewrite the rough Bergomi model on forward variance curve form. We thus present the below result:

Theorem 6.1.1. *The rough Bergomi model can be equivalently written under the pricing measure Q as*

$$dS_t = (r_t - q_t)S_t dt + S_t \sqrt{\xi_t(t)} dW_{2,t} \quad (6.1)$$

and

$$d\xi_t(u) = \xi_t(u) \tilde{\eta}(u-t)^{-\gamma} dW_{1,t}. \quad (6.2)$$

for all $u > t$, where $\tilde{\eta} := \eta \sqrt{2H}$ and ξ_0 and S_0 are observable at time zero.

Proof. Let $u > t$. First see that

$$\xi_t(u) = E_t^Q(\sigma_u^2) \quad (6.3)$$

$$= E_t^Q\left(\xi_0(u)\mathcal{E}\left(\eta\tilde{W}_0(u)\right)\right) \quad (6.4)$$

$$= E_t^Q\left(\xi_0(u)e^{-\frac{1}{2}\eta^2 u^{2H} + \eta\tilde{W}_0(u)}\right) \quad (6.5)$$

$$= \xi_0(u)e^{-\frac{1}{2}\eta^2 u^{2H}} E_t^Q\left(e^{\eta\tilde{W}_0(u)}\right). \quad (6.6)$$

Now take note of the following decomposition

$$\eta\tilde{W}_0(u) = \eta\sqrt{2H} \int_0^u (u-s)^{-\gamma} dW_{1,s} \quad (6.7)$$

$$= \tilde{\eta} \int_0^t (u-s)^{-\gamma} dW_{1,s} + \tilde{\eta} \int_t^u (u-s)^{-\gamma} dW_{1,s} \quad (6.8)$$

where the first term is measurable with respect to \mathcal{F}_t and the last term is independent of \mathcal{F}_t and normally distributed with mean zero and variance

$$\text{Var}\left(\tilde{\eta} \int_t^u (u-s)^{-\gamma} dW_{1,s}\right) = \tilde{\eta}^2 \int_t^u (u-s)^{-2\gamma} ds = \tilde{\eta}^2 \frac{(u-t)^{2H}}{2H} = \eta^2 (u-t)^{2H}. \quad (6.9)$$

We therefore get

$$\xi_t(u) = \xi_0(u)e^{-\frac{1}{2}\eta^2 u^{2H}} E_t^Q\left(e^{\eta\tilde{W}_0(u)}\right) \quad (6.10)$$

$$= \xi_0(u)e^{-\frac{1}{2}\eta^2 u^{2H} + \frac{1}{2}\eta^2(u-t)^{2H}} e^{\tilde{\eta} \int_0^t (u-s)^{-\gamma} dW_{1,s}} \quad (6.11)$$

$$= \xi_0(u)e^{\frac{1}{2}\eta^2((u-t)^{2H} - u^{2H})} e^{\tilde{\eta} \int_0^t (u-s)^{-\gamma} dW_{1,s}} \quad (6.12)$$

$$= f_u(t)e^{\tilde{Y}_t(u)} \quad (6.13)$$

where we have defined the deterministic function $f_u(t) := \xi_0(u)e^{\frac{1}{2}\eta^2((u-t)^{2H} - u^{2H})}$ and the Ito process $\tilde{Y}_t(u) := \tilde{\eta} \int_0^t (u-s)^{-\gamma} dW_{1,s}$.

Note that $d\tilde{Y}_t(u) = \tilde{\eta}(u-t)^{-\gamma} dW_{1,t}$ such that an application of Ito gives

$$\begin{aligned} d\xi_t(u) &= \frac{\partial f_u(t)}{\partial t} e^{\tilde{Y}_t(u)} + f_u(t) e^{\tilde{Y}_t(u)} d\tilde{Y}_t(u) + \frac{1}{2} f_u(t) e^{\tilde{Y}_t(u)} (d\tilde{Y}_t(u))^2 \\ &= -\frac{1}{2}\tilde{\eta}^2(u-t)^{2H-1} \xi_t(u) dt + \xi_t(u) \tilde{\eta}(u-t)^{-\gamma} dW_{1,t} + \frac{1}{2} \xi_t(u) \tilde{\eta}^2(u-t)^{-2\gamma} dt \\ &= \xi_t(u) \tilde{\eta}(u-t)^{-\gamma} dW_{1,t} \end{aligned}$$

where we have also used that $\gamma = \frac{1}{2} - H$.

Note also that we recover the variance process at time t as the limit $\lim_{u \rightarrow t^+} \xi_t(u) = \xi_t(t)$. Thus it is an equivalent way of writing the rough Bergomi model under the pricing measure Q . \square

We now turn our attention to the question of finding the dynamics of the time $t > 0$ price of some simple claim with time $T > t$ payoff $g(S_T)$ with g being the payoff function. In a model driven only by stochastic differential equations (SDE's) one can use the Markov property to conclude that the price is directly a function of only the state variables. However, in our case the volatility process

is driven by a fractional Brownian motion and so does not satisfy a SDE. In particular it is not Markov¹ and so the price cannot generally be written as a function of (t, S_t, σ_t) .

However, using the equivalent Q -model specification in theorem 6.1.1 we see that the model is instead Markov in (S_t, ξ_t) . We therefore conclude

$$E^Q(g(S_T)|\mathcal{F}_t) = E^Q(g(S_T)|S_t, \xi_t) \quad (6.14)$$

and thus the price of the derivative can be written as

$$P(t, S_t, \xi_t) := e^{-\int_t^T r(s) - q(s) ds} E^Q(g(S_T)|\mathcal{F}_t). \quad (6.15)$$

where P is a *function* of time t and spot S_t but a *functional* of the forward variance curve ξ_t .²

Remark. *Keep in mind that P is, among other things, a functional of the forward variance curve ξ_t which is infinite-dimensional. However, since all the forward variances are driven by the same Brownian motion one might think that we could reduce the dimensionality of P further. In particular one might expect to be able to reduce it to a function of the form $P(t, S_t, \xi_t(u))$ where $u > T$ is some fixed forward variance maturity.*

That is not possible for the following reason:

From the proof of theorem 6.1.1 we recall the formulation

$$\xi_t(u) = f_u(t) e^{\tilde{Y}_t(u)} \quad (6.16)$$

where

$$\tilde{Y}_t(u) = \tilde{\eta} \int_0^t (u-s)^{-\gamma} dW_{1,s} \quad (6.17)$$

and $f_u(t)$ is some specific deterministic function.

*As a thought experiment say we were able to write $\tilde{Y}_t(u) = Z_t \cdot h(u, t)$ where h is a deterministic function of u and t and $(Z_t)_{t \geq 0}$ is an adapted stochastic process common for all forward variances. Then knowing $\xi_t(u')$ for some fixed u' we would know the value of Z_t and thus be able to infer the value of the entire forward variance curve ξ_t . In that case we would indeed be able to reduce P to a function of just two state variables. Unfortunately, due to the form of the kernel $(u-s)^{-\gamma}$ in the stochastic integral in (6.17) we are unable to separate u from the stochastic factor W_1 .*³

□

As the arguments that follow apply in a general forward variance curve framework we will now expand our analysis to this more general class of models. Under this class of models we assume that the combined dynamic system of the stock price and all the forward variances follow a general SDE.⁴ With this assumption there naturally are no jumps and furthermore the covariations will be of the following very general form

$$(d\log(S_t))^2 = \sigma^2(t, S, \xi) dt \quad (6.18)$$

$$d\xi_t(u)d\log(S_t) = \mu(t, u, S, \xi) dt \quad (6.19)$$

$$d\xi_t(u)d\xi_t(u') = \nu(t, u, u', S, \xi) dt \quad (6.20)$$

¹See (Hu 2003) for a rigorous proof.

²We refer the reader to section B.4 in the appendix for a brief explanation of functionals.

³In fact it is only in the non-rough case of $H = \frac{1}{2}$, i.e. $\gamma = 0$, that we can reduce the dimensionality of P further.

⁴We remark that the forward variances must be martingales under the pricing measure to avoid arbitrage. See section B.5 in the appendix for why this must be. Thus we know more specifically $d\xi_t(u) = \dots dW_t$ where W_t is a Q -Brownian motion.

for $u, u' > t$ where σ, μ and ν are functions. However as the models we will refer to are such that μ and ν does not depend on the stock we will make exactly this assumption. That is we further assume $\mu(t, u, S, \xi) = \mu(t, u, \xi)$ and $\nu(t, u, u', S, \xi) = \nu(t, u, u', \xi)$.

As before we still have that the price of a simple claim is of the form $P(t, S, \xi)$. No more assumptions are needed at this stage.

Note briefly that the rough Bergomi model is recovered as

$$\sigma^2(t, S, \xi) = \xi(t) \quad (6.21)$$

$$\mu(t, u, \xi) = \sqrt{\xi(t)}\xi(u)\tilde{\eta}\rho(u-t)^{-\gamma} \quad (6.22)$$

$$\nu(t, u, u', \xi) = \xi(u)\xi(u')\tilde{\eta}^2[(u-t)(u'-t)]^{-\gamma}. \quad (6.23)$$

Consider now an unhedged long position in the derivative financed via a short position in the risk-free asset. As the forward variance curve is infinite dimensional we cannot use the usual Ito's lemma to find the dynamics of P . However, a more general version that can handle the infinite dimensional forward variance curve does exist and yields the following result:⁵

$$dP(t, S, \xi) = \frac{\partial P}{\partial t}dt + \frac{\partial P}{\partial S}dS + \int_t^T \frac{\delta P}{\delta \xi(u)}d\xi(u)du \quad (6.24)$$

$$+ \frac{1}{2} \frac{\partial^2 P}{\partial S^2}(dS)^2 + \frac{1}{2} \int_t^T \int_t^T \frac{\delta^2 P}{\delta \xi(u)\delta \xi(u')}d\xi(u)d\xi(u')dudu' \quad (6.25)$$

$$+ \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)}dSd\xi(u)du. \quad (6.26)$$

We remark that the Ito expansion above looks remarkably much as if one had used the standard Ito formula. However, expressions such as $\frac{\delta P}{\delta \xi}$ are so-called functional derivatives since the ordinary derivative of a function (or functional rather), here P , with respect to another function, here ξ_t , is not well-defined. We provide a brief note on the subject in the appendix - see section B.4.

In order to hedge the derivative we first need to get rid of the stochastic terms, i.e. the terms involving dS and $d\xi(u)$. Thus we first short $\frac{\partial P}{\partial S}$ shares and invest the proceeds in the risk-free asset. In order to get rid of the term involving $d\xi_t(u)$ we note that one can construct a trading strategy in variance swaps that costs zero and pays a dividend of exactly $d\xi_t(u)$. We refer to such a constructed contract as a *forward variance swap*. That is, one can actually trade the forward variances and that at zero cost. We provide a proof of this construction in the appendix - see section B.5. Thus to sum up, we also go short $\frac{\delta P}{\delta \xi(u)}$ forward variance swaps for each $u \in (t, T]$. This costs nothing and provides a per period PnL of exactly $-\int_t^T \frac{\delta P}{\delta \xi(u)}d\xi(u)du$.

Keeping track of dividend payments from our stock position and as well as the position in the risk-free asset we now have the following PnL:

$$\text{PnL} = \frac{\partial P}{\partial t}dt + \frac{1}{2} \frac{\partial^2 P}{\partial S^2}(dS)^2 + \frac{1}{2} \int_t^T \int_t^T \frac{\delta^2 P}{\delta \xi(u)\delta \xi(u')}d\xi(u)d\xi(u')dudu' \quad (6.27)$$

$$+ \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)}dSd\xi(u)du - r(t)Pdt + (r(t) - q(t))S \frac{dP}{dS}dt. \quad (6.28)$$

Note that under our continuous semi-martingale assumptions the PnL above is locally deterministic. That is, it only involves terms of the form ' dt '. As a consequence, to avoid arbitrage the drift must

⁵A good reference for stochastic differential equations in infinite dimensions is (Gawarecki & Mandrekar 2011).

equal zero. Thus we get the infinite-dimensional partial differential equation (PDE)

$$\frac{\partial P}{\partial t} + (r(t) - q(t))S \frac{dP}{dS} + \frac{1}{2} \frac{\partial^2 P}{\partial S^2} S^2 \sigma^2(t, S, \xi) + \frac{1}{2} \int_t^T \int_t^T \frac{\delta^2 P}{\delta \xi(u) \delta \xi(u')} \nu(t, u, u', \xi) dudu' \quad (6.29)$$

$$+ \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} \mu(t, u, \xi) du = r(t)P \quad (6.30)$$

and that with terminal condition $P(T, S, \xi) = g(S)$.⁶

Say now we use a model with potentially incorrect covariation functions σ, μ, ν . That is we let our pricing function P satisfy the PDE (6.29)-(6.30) but where the functions σ, μ, ν do not match the actual covariations as in (6.18) - (6.20). In that case the PnL equation (6.27)-(6.28) will still hold and plugging the PDE satisfied by P back into (6.27)-(6.28) we get the following PnL equation:

$$\text{PnL} = \frac{S^2}{2} \frac{\partial^2 P}{\partial S^2} \left(\left(\frac{dS}{S} \right)^2 - \sigma^2(t, S, \xi) dt \right) \quad (6.31)$$

$$+ \frac{1}{2} \int_t^T \int_t^T \frac{\delta^2 P}{\delta \xi(u) \delta \xi(u')} (d\xi(u) d\xi(u') - \nu(t, u, u', \xi) dt) dudu' \quad (6.32)$$

$$+ \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} S \left(\frac{dS}{S} d\xi(u) - \mu(t, u, \xi) dt \right) du. \quad (6.33)$$

We can think of P as the pricing formula we use for hedging purposes. The PnL in equation (6.31)-(6.33) is then the PnL we earn hedging with the potentially incorrect pricing function P . Thus (6.31)-(6.33) is perhaps best described as an accounting relation for our PnL. This interpretation will be useful when we later consider trading in a misspecified market. With respect to terminology we will refer to the term (6.31) as the spot-gamma PnL, to (6.32) as the volatility-gamma PnL and to (6.33) as the spot/volatility PnL. Also, and as expected, if P is the correct pricing function all these terms become zero and we hedge the derivative perfectly.

6.2 The Skew-Stickiness-Ratio

In this section we consider the so-called skew-stickiness-ratio (abbreviated SSR). For that purpose note that we can always write the price of a vanilla option with expiry $T > 0$ and strike K , $k = \log(K/F_T)$ being log-moneyness, as $P_{BS}(t, S, \sigma_{BS}(k, T))$ where $P_{BS}(t, S, \sigma)$ is the Black-Scholes formula for an expiry $T > 0$ at-the-money vanilla option with time-to-expiry $T - t$, current stock price S and a Black-Scholes volatility of σ , $\sigma_{BS}(k, T)$ being the implied volatility of the option. The delta-hedge at time t can then be computed as

$$\frac{dP_{BS}(t, S_t, \sigma_{BS}(k, \tau))}{dS} = \frac{\partial P_{BS}}{\partial S} + \frac{\partial P_{BS}}{\partial \sigma} \cdot \frac{d\sigma_{BS}(k, \tau)}{dS}. \quad (6.34)$$

As the Black-Scholes greeks $\frac{\partial P_{BS}}{\partial S}$ and $\frac{\partial P_{BS}}{\partial \sigma}$ are easily computed given the observed implied volatility what we see is that the main component in computing the correct delta-hedge is knowing the relation between implied volatility and the stock price, i.e. $\frac{d\sigma_{BS}(k, \tau)}{dS}$. Various ad hoc rules for that relation could be mentioned here. One example is the so-called sticky-strike rule which postulates that implied volatility is only a function of strike. The implication is that the implied volatility does not change when the spot moves. This is of course a highly unrealistic assumption and essentially an attempt of maintaining a Black-Scholes like hedging framework. Another commonly known rule is

⁶We remark that since there is an infinite number of state variables solving this PDE is highly infeasible, both in a general setting as well as for the rough Bergomi model.

the sticky-moneyness rule that postulates that implied volatility is only a function of moneyness. The implication here is that implied volatility moves locally one-to-one with skew. However, generally speaking many such relations can be proposed. Thus to be more rigorous on exactly how implied volatility and spot vary together we need to introduce a new quantity that captures this dynamic for a given model. To this end we define the expiry $T > 0$ SSR as

$$R_T := \frac{1}{\psi(T)} \frac{E(d \ln S d\hat{\sigma}_{F_T})}{E((d \ln S)^2)} \quad (6.35)$$

where $\hat{\sigma}_{F_T}$ is the at-the-money ($k = 0$) implied volatility for an expiry T vanilla option.⁷ Note that the expectation operator $E(\cdot)$ is irrelevant if both $\ln(S)$ and $\hat{\sigma}_{F_T}$ are continuous semi-martingales.

Intuitively, we can think of the second factor in the above as the regression coefficient when regressing a move in the at-the-money implied volatility on a move in the log-spot. The SSR thus quantifies how much the at-the-money implied volatility moves when the underlying moves and that relative to the at-the-money skew. Different models will generate different dynamics for the smile and so we should generally expect them to produce different values of SSR. Recalling the ad hoc rules we see heuristically that $R_T = 1$ for the sticky-strike rule as the at-the-money implied volatility slides one-to-one with skew along the smile as the spot moves. For the sticky-moneyness rule the smile is fixed in terms of moneyness and thus the at-the-money implied volatility does not move under this rule thus implying $R_T = 0$.

A number of papers have presented empirical estimations of the SSR for various indexes.⁸ In (Bergomi 2009) it was estimated on the Eurostoxx50 index that $R_T \in [1, 2]$ for both short and long term options. However, mostly the SSR was found to be close to 1.5 being a little higher for short expiry options but still moderately below 2. These results were also verified for the S&P 500 in (Bergomi 2016) - see section 9.7.

Because of the importance of the SSR in summarizing the dynamic properties of a given model and thus its implications for hedging we now give a brief overview of the values produced by different models. In particular we want to see how well they can be made to fit the empirical results just explained. As is claimed in (Bergomi 2016) local volatility models generally produce $R_T \rightarrow 2$ for $T \rightarrow 0$. It is also found that a local volatility model calibrated to a typical equity smile will display SSR values above 2 for longer expiries. In (Bergomi 2016) it is furthermore claimed that jump diffusion models with independent and stationary increments for the log-spot produce $R_T = 0$. Finally we can consider the large class of stochastic volatility models. This class contains well-known models such as Heston as well as the n -factor Bergomi model explored in (Bergomi 2016).⁹ Importantly it also contains the rough Bergomi model we consider in this thesis. It is useful to note that since a stochastic volatility model models the spot volatility $(\sigma_t)_{t \geq 0}$ we can usually rephrase such a model on forward variance curve form by computing the dynamics of $\xi_t(u) := E^Q(\sigma_u^2 | \mathcal{F}_t)$. We may therefore as well just consider the entire class of forward variance curve models. We will however restrict our attention to models where the forward variances are continuous semi-martingales.

In (Bergomi & Guyon 2011) Lorenzo Bergomi and Julien Guyon derives an expansion formula

⁷We use the extra subscript F_T to stress that since we consider the ATMF implied volatility the strike is not fixed across time. Specifically $\hat{\sigma}_{F_T}$ is the implied volatility of a maturity T , strike F_T vanilla option where F_T is the forward price.

⁸It is important to mention that the estimates we refer to here are of the so-called *realized* SSR. The realized SSR is defined as in (6.35) except the skew is computed from option prices, i.e. observed under the risk neutral measure, whereas the second factor in the expression is estimated under the historical probability measure. If the model used by the market implies the correct covariation between the spot and the at-the-money-forward implied volatility as well as the correct quadratic variation of the stock price then the realized version of the SSR and the definition in (6.35) will be the same.

⁹Go to page 99 in the appendix to recap the definition of these models.

for implied volatilities in a general class of multi-factor diffusive stochastic volatility models. For convenience we state their result in section B.6 in the appendix. With a model on forward variance curve form one can use the Bergomi-Guyon expansion formula to derive the following approximate formula for skew

$$\psi(T) = \frac{\int_0^T \int_\tau^T \mu(\tau, u, \xi_0) du d\tau}{2\sqrt{T} \left(\int_0^T \xi_0(\tau) d\tau \right)^{3/2}} \quad (6.36)$$

and the following for the SSR

$$R_T = \frac{\int_0^T \xi_0(\tau) d\tau}{T \xi_0(0)} \frac{T \int_0^T \mu(0, u, \xi_0) du}{\int_0^T \int_\tau^T \mu(\tau, u, \xi_0) du d\tau}. \quad (6.37)$$

A derivation of these result as well as the remaining results we will present in this subsection can be found in section B.7 in the appendix. Importantly from (6.37) we see that while the forward variance curve also enters into the formula, the SSR is largely a function of how the spot/variance covariance function μ is modelled. Furthermore, using the above expressions and assuming in addition that $u \mapsto \mu(t, u, \xi)$ as well as $u \mapsto \xi_t(u)$ are well-defined as well as sufficiently differentiable in the point $u = t$ we argue in the appendix that $R_0 := \lim_{T \rightarrow 0} R_T = 2$ and also that $\psi(0) := \lim_{T \rightarrow 0} \psi(T)$ is constant. While these results will hold for most conventional stochastic volatility models, including Heston, it will not hold for the rough Bergomi model as the power-law kernel $(u - t)^{-\gamma}$ is not even well defined in $u = t$. The first result suggests that most stochastic volatility models will be unable to display a SSR of 1.5 for short expiries and the second that they will not be able to replicate the exploding skew observed in practise. It thus seems that any model that wishes to replicate these central empirical facts need the spot/variance covariance function to have some sort of singularity at the origin $u = t$.

Assume now that the forward variance curve to an approximation is flat and that $\mu(t, u, \xi_0)$ only depends on u and t through $u - t$. Let us then abuse notation and write $\mu(u - t)$ instead of $\mu(t, u, \xi_0)$ as well as write ξ_0 to refer both to the forward-variance curve *function* as well as the fixed *value* it takes. Assuming further that $\mu(t)$ decays monotonically towards zero for $t \rightarrow \infty$ you can derive the following approximative formulas

$$\psi(T) = \frac{1}{2\xi_0^{3/2} T^2} \int_0^T g(\tau) d\tau \quad (6.38)$$

$$R_T = \frac{g(T)}{\frac{1}{T} \int_0^T g(\tau) d\tau} \quad (6.39)$$

where we have defined $g(\tau) := \int_0^\tau \mu(t) dt$. Using the above approximate expression for the SSR one can argue that $R_T \in [1, 2]$ thus at least for all such models constraining the SSR to be roughly around the empirical observed values of approximately 1.5. However this bound is quite wide and it is also important to note that this result is only true up to a flat forward variance curve. Thus a given model may in practise produce values outside this range.

With the above expressions in mind we now wish to say a little bit more about how skew and stickiness behaves for $T \rightarrow \infty$. To this end we maintain the assumption of a time-homogeneous spot/variance covariance function and further assume $\mu(t) \sim t^{-\gamma}$ for $t \rightarrow \infty$.¹⁰

As we argue in the appendix this identifies two types of models depending on the value of γ :

¹⁰Here the symbol ' \sim ' simply means that the ratio of the left- and right-hand-side converge to a constant, i.e. are roughly proportional for large t .

Type I: If $\gamma > 1$ then $\psi(T) \sim T^{-1}$ for large T and $R_\infty = 1$.

Type II: If $\gamma < 1$ then $\psi(T) \sim T^{-\gamma}$ for large T and $R_\infty = 2 - \gamma$.

Here we use the notation $R_\infty := \lim_{T \rightarrow \infty} R_T$.

We conclude, under the assumptions introduced so far, that the rate of decay of skew and the value of the SSR, both for long expiries, are really two sides of the same coin and are both related to the rate of decay of the spot/variance covariance function $\mu(t)$. Also, while the classification above only allows spot-variance covariance functions $\mu(t)$ that decay asymptotically as a power-law, one can argue that if $\mu(t)$ instead has an exponential form then it exhibits exactly the properties of a type I model. In particular models such as Heston and the n -factor Bergomi model have exponential kernels. In light of these results we can then conclude that such models should have a term structure of SSR that starts at 2 in the origin and decays to 1 in the limit $T \rightarrow \infty$ and term structure of skew that starts from a constant value at the origin and then decays as T^{-1} in the limit $T \rightarrow \infty$. Both these properties are at odds with the empirical observations thus highlighting the need for an entirely new generation of pricing models.¹¹

We note that $\mu(t) = \sqrt{\xi_0(0)}\xi_0(t)\tilde{\eta}t^{-\gamma}$ for the rough Bergomi model and thus, to the approximation of a flat forward variance curve, it is a type II model with $R_\infty = 2 - \gamma = \frac{3}{2} + H$. In fact, assuming a flat forward variance curve, equation (6.39) will even suggest $R_T = \frac{3}{2} + H$ for all $T > 0$. To see this we compute

$$g(T) = \int_0^T \sqrt{\xi_0(0)}\xi_0(t)\tilde{\eta}t^{-\gamma}dt = \xi_0(0)^{3/2}\tilde{\eta} \int_0^T t^{-\gamma}dt = \xi_0(0)^{3/2}\tilde{\eta}(1-\gamma)^{-1}T^{1-\gamma} \quad (6.40)$$

such that

$$\frac{1}{T} \int_0^T g(t)dt = \frac{1}{T}\xi_0(0)^{3/2}\tilde{\eta}(1-\gamma)^{-1} \int_0^T t^{1-\gamma}dt = \xi_0(0)^{3/2}\tilde{\eta}(1-\gamma)^{-1}(2-\gamma)^{-1}T^{1-\gamma} \quad (6.41)$$

and inserting these results into (6.39) we get the suggestion $R_T = 2 - \gamma = \frac{3}{2} + H$. Since the calibration from chapter 5 suggested $H \approx 0.1$ we should expect a calibrated rough Bergomi mode to not only fit the power-law decay of skew but also to produce SSR values close to 1.5 for all expiries as is empirically observed.

To confirm the approximative results we in figure 6.1 (see the next page) show the term structure of the SSR for the rough Bergomi model computed numerically.¹² Here we roughly confirm the approximative SSR result of $R_T = \frac{3}{2} + H$ for various choices of H . In the case of $H = 0.1$ we also tried using an increasing forward variance curve $\xi_0(u) = 0.065 - 0.03e^{-0.5u}$ and a decreasing one $\xi_0(u) = 0.02 + 0.05e^{-5u}$. Here we confirm that the SSR values indeed do deviate somewhat from the case of flat curve. However interestingly enough the effects of a non-flat forward variance curve seems to disappear as $T \rightarrow 0$. It also seems that the SSR values for $H = 0.1$, $H = 0.2$ and $H = 0.3$ are closest to the expected values of $\frac{3}{2} + H$ when T is close to zero. In other words it seems second

¹¹We do remark that Lorenzo Bergomi in (Bergomi 2016) argues that the n -factor Bergomi model with $n = 2$ stochastic factors driving the forward variance curve is actually able to display SSR values of around 1.5 for most relevant expiries thus essentially building an approximate type II model from a type I model. However with 2 factors the model contains 8 parameters in total which is to be compared to the alternative choice of the rough Bergomi model with its 3 parameters. In (Gatheral 2014) Jim Gatheral also claims that this model is overparameterised and calibrated parameters tend to display a high correlation between the 2 stochastic factors driving the forward variance curve.

¹²We explain our numerical method in section B.7.4 in the appendix.

order effects not accounted for in (6.39) take hold for moderately long expiries as the SSR values drift somewhat away from $\frac{3}{2} + H$. As the rough Bergomi model is a type II model we should however still expect the long term limit to also be $\frac{3}{2} + H$. Looking at the curves for $H = 0.2$, $H = 0.3$ this is difficult to see and so we are unable to properly confirm that theoretical result though we still think it likely to be true based on the theoretical arguments we provide in the appendix for type II models.

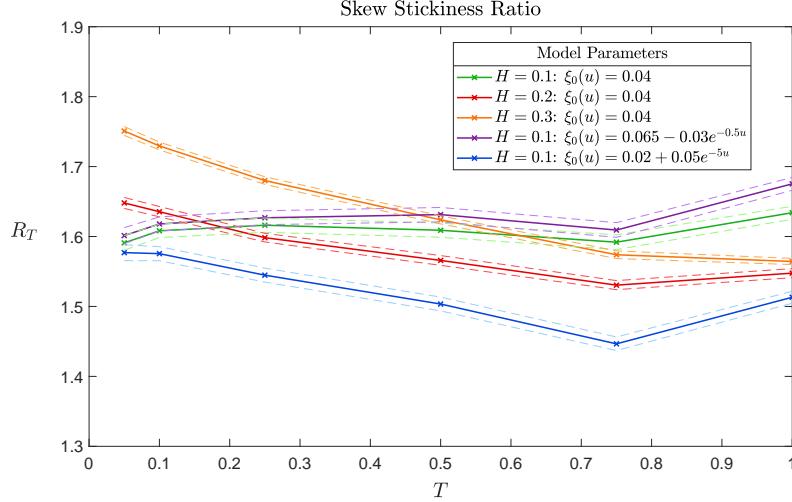


Figure 6.1: The term structure of the SSR of the rough Bergomi model under various assumptions. The following parameters are fixed: $\rho = -0.9$, $\eta = 1.9$, $S_0 = 100$, $r = 0.05$, $q = 0.02$. Each estimate is based on the mean of 100 independent estimations of R_T using the method from section B.7.4 in the appendix. Dashed lines are 95 % confidence intervals computed from these 100 samples by bootstrapping with 100.000 samples.

6.3 Trading the Skew-Stickiness-Ratio

In light of the PnL accounting equation (6.31) - (6.33) and with the knowledge of the concept of the SSR we are ready to consider arbitraging it in a misspecified market. Assume therefore that both the actual model that governs the dynamics of the spot price and the forward variances and the model used by the market to price options are within the set-up from section 6.1. We will refer to the latter as the *market* or *implied* model. In this setting we will write superscript '*a*' if an object is from the actual model and '*i*' if it is from the market model. As an example $d\xi_t^i(u)$ refers to the dynamics of the expiry u forward variance as implied by the market model and $d\xi_t^a(u)$ to the dynamics that the forward variance follows under the actual model. We will consider mark-to-market accounting and thus we will let P be the pricing function that results from using the market model. Thus P will be the solution to the PDE in (6.29) - (6.30) where $\mu(t, u, \xi_t)dt = d\log(S_t^i)d\xi_t^i(u)$ and similarly for the functions ν and σ . Hedging with the market model then exactly results in a per period PnL as given by the terms (6.31) - (6.33) where all dynamic terms of the form $d(\cdot)$ should have superscript '*a*'.

There are three PnL terms to keep track of. We will argue that the spot/volatility PnL is realized as the difference between the realized SSR (to be rigorously defined) and the SSR under the market model. In light of the relationship $R_T = \frac{3}{2} + H$ for the rough Bergomi model this is what we are interested in. However to isolate that term one needs to appropriately eliminate the other two PnL terms. As all the models we consider take the current forward variance curve as input

the spot-gamma PnL term should in theory be zero as any reasonable market model will have $\sigma^2(t, S_t, \xi_t) = \xi_t(t) = \lim_{u \rightarrow t^+} \xi_t(u)$ which is computable knowing ξ_t . In practise we cannot observe the entire forward variance curve or we can only observe it with error. Some sort of hedging of the spot-gamma term may therefore be necessary eliminate or minimize this particular term in a real world setting. Since we will consider the model in continuous time and with full knowledge of the forward variance curve we though for our purpose assume that the spot-gamma PnL is zero. This only leaves us to handle the volatility-gamma PnL properly. We will however delegate that part to future research as there is plenty to say about the spot-volatility PnL.

We will now argue that the spot-volatility PnL realizes a PnL proportional to the difference between realized and implied SSR. To arrive at something that is analytical we need to use a decent amount of approximate expressions. While we will present numerical computations checking some of the approximations we generally leave large scale numerical tests for future research.

First, assuming that both the stock and forward variances are continuous semi-martingales we may use a result we prove on page 93 in the appendix to conclude the general relation

$$\int_t^T \frac{dS_t}{S_t} d\xi_t(u) du dt = 2(T-t)\hat{\sigma}_T^{VS}(t) d\ln S_t d\hat{\sigma}_T^{VS}(t). \quad (6.42)$$

Two versions of the above relation will then hold. One under the market model where we add superscripts '*i*' and one under the actual model where we add superscripts '*a*'. Note also that since the current forward variance curve ξ_t is an input to the models and is observable both models will agree on its value. In particular they will agree on the implied variance swap volatility $\hat{\sigma}_T^{VS}(t)$. Thus it is really only the $d(\cdot)$ components in (6.42) that needs superscripts.

Secondly, according to the Bergomi-Guyon expansion formula the implied variance swap volatility will roughly equal the at-the-money implied volatility. Thus we should expect $d\hat{\sigma}_T^{VS} \approx d\hat{\sigma}_{F_T}$.

Thirdly we assume the following hypothesis:

$$H^* : \text{For a fixed } t \in [0, T) \text{ the mapping } u \mapsto \frac{\partial \delta P}{\partial S \delta \xi(u)} \text{ is constant.} \quad (6.43)$$

Let us denote this most likely stochastic but certainly measurable 'constant' value by $\tilde{C}(t)$. We later check how valid the hypothesis is in the special case of a vanilla option in the rough Bergomi model.

Keeping all these results, approximations and H^* in mind we are now ready to rewrite the spot-

volatility PnL:

$$\int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} S_t \left(\frac{dS_t^a}{S_t} d\xi_t^a(u) - \mu(t, u, \xi_t) dt \right) du \quad (6.44)$$

$$= S_t \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} \frac{dS_t^a}{S_t} d\xi_t^a(u) du - S_t \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} \frac{dS_t^i}{S_t} d\xi_t^i(u) du \quad (6.45)$$

$$\approx S_t \cdot \tilde{C}(t) \int_t^T \frac{dS_t^a}{S_t} d\xi_t^a(u) du - S_t \cdot \tilde{C}(t) \int_t^T \frac{dS_t^i}{S_t} d\xi_t^i(u) du \quad (6.46)$$

$$= 2S_t \tilde{C}(t)(T-t) \hat{\sigma}_T^{VS}(t) d \ln S_t^a d \hat{\sigma}_T^{VS,a}(t) - 2S_t \tilde{C}(t)(T-t) \hat{\sigma}_T^{VS}(t) d \ln S_t^i d \hat{\sigma}_T^{VS,i}(t) \quad (6.47)$$

$$= 2S_t \tilde{C}(t)(T-t) \hat{\sigma}_T^{VS}(t) \left(d \ln S_t^a d \hat{\sigma}_T^{VS,a}(t) - d \ln S_t^i d \hat{\sigma}_T^{VS,i}(t) \right) \quad (6.48)$$

$$\approx 2S_t \tilde{C}(t)(T-t) \hat{\sigma}_T^{VS}(t) \psi^i(T-t) \xi_t(t) \left(\frac{d \ln S_t^a d \hat{\sigma}_{FT}^a(t)}{\psi^i(T-t) \xi_t(t)} - \frac{d \ln S_t^i d \hat{\sigma}_{FT}^i(t)}{\psi^i(T-t) \xi_t(t)} \right) \quad (6.49)$$

$$= 2S_t \tilde{C}(t)(T-t) \hat{\sigma}_T^{VS}(t) \psi^i(T-t) \xi_t(t) (R_{T-t}^r - R_{T-t}^i) \quad (6.50)$$

where we have also defined the *realized* SSR more generally as

$$R_T^r := \frac{1}{\psi^i(T)} \frac{E[d \ln S^a d \hat{\sigma}_{FT}^a]}{E[(d \ln S^a)^2]}. \quad (6.51)$$

Note that this is *not* the same as the SSR under the actual model as we are dividing by the skew under the market implied model and not under the actual model. This is also why we are not using the notation R_T^a . However if we choose an expiry where we have $\psi^a(T) = \psi^i(T)$ then we indeed get $R_T^r = R_T^a$. Under our hypothesis and for the case of the rough Bergomi model choosing such an expiry should result in a per period spot/volatility PnL that is roughly proportional to the difference in Hurst exponents:

$$R_{T-t}^a - R_{T-t}^i = 2 - \gamma^a - (2 - \gamma^i) = \gamma^i - \gamma^a = H^a - H^i. \quad (6.52)$$

Regardless, as is revealed in (6.50) the nature of the spot/volatility PnL is more generally that of betting on the difference between realized and implied SSR. Thus we will continue to consider more generally the problem of realizing a total PnL that is proportional to the difference in realized and implied SSR rather than explicitly the difference in Hurst exponents. This is particularly so in light of the somewhat ambiguous results in figure 6.1 revealing that while H does have a measurable impact on the SSR it is only to a coarse approximation true that we exactly have $R_T = \frac{3}{2} + H$, in particular for longer expiries.

Let us investigate our hypothesis that the sensitivity of delta with respect to each point on the forward variance curve, i.e. $u \mapsto \frac{\partial \delta P}{\partial S \delta \xi(u)}$, is roughly constant along the forward variance curve expiries. In section B.4.2 in the appendix we propose a method to compute this function numerically. For the purpose of the numerical computations we now let P be the price of a vanilla option. In figure 6.2 we then keep the expiry at $T = \frac{1}{2}$ and plot the functional derivative under a specific rough Bergomi model for a few choices of log-moneyness k .

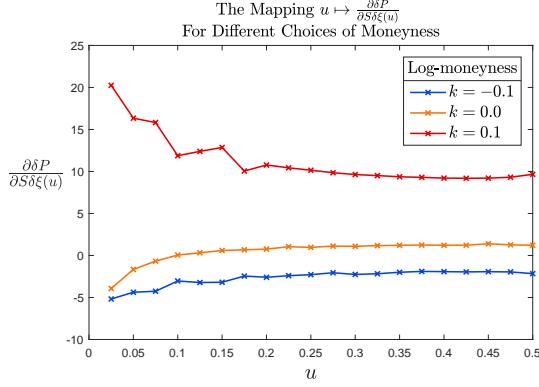


Figure 6.2: Plot of the function $u \mapsto \frac{\partial \delta P}{\partial S \delta \xi(u)}$ for different choices of log-moneyness k where P is the price of a vanilla option under the rough Bergomi model. The fixed parameters are $H = 0.1, \eta = 1.9, \rho = -0.9, r = 0.05, q = 0.02, S_0 = 100$ and $\xi_0(t) = 0.04$ for all t . The expiries of the options are fixed at $T = \frac{1}{2}$.

From the figure we see that $u \mapsto \frac{\partial \delta P}{\partial S \delta \xi(u)}$ generally is not constant. While this violates H^* for the rough Bergomi model we still hope the result linking (6.44) to (6.50) will hold to an approximation if we instead let

$$\tilde{C}(t) = \frac{1}{T-t} \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} du \quad (6.53)$$

be the average value. Let us thus for now assume that the equations (6.44) - (6.50) approximately holds with this choice of \tilde{C} .

Say now we have a good prediction on the sign of $R_{T-t}^r - R_{T-t}^i$, for instance because we think the market uses an incorrect Hurst exponent in rough Bergomi model. In order to decide on which side of the trade to position ourselves all that is left to do is somehow reason about the sign of the pre-factor which under the choice from equation (6.53) can be rewritten as

$$2S_t \int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} du \cdot \hat{\sigma}_T^{VS}(t) \psi^i(T-t) \xi_t(t). \quad (6.54)$$

For equity smiles we typically observe $\psi^i(T) < 0$ for all expiries and thus we are only left with reasoning about the sign of $\int_t^T \frac{\partial \delta P}{\partial S \delta \xi(u)} du$. For that purpose we in figure 6.3 plot the integral $\int_0^T \frac{\partial \delta P}{\partial S \delta \xi(u)} du$ across moneyness.

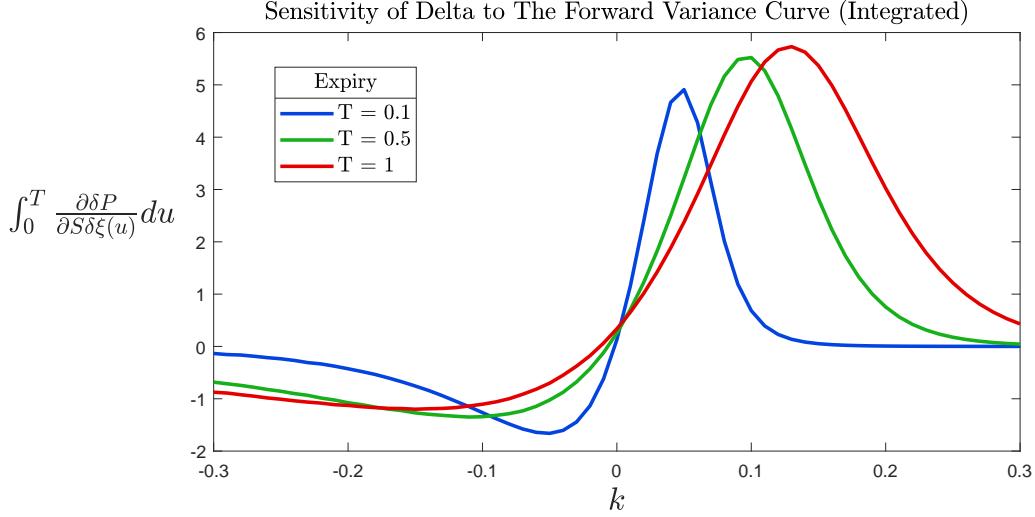


Figure 6.3: Plot of $\int_0^T \frac{\partial \delta P}{\partial S \partial \xi(u)} du$ across moneyness where P is the price of a vanilla option under the rough Bergomi model. The fixed parameters are $H = 0.1, \eta = 1.9, \rho = -0.9, r = 0.05, q = 0.02, S_0 = 100$ and $\xi_0(t) = 0.04$ for all t .

The figure suggests that $\int_0^T \frac{\partial \delta P}{\partial S \partial \xi(u)} du$ is negative for values of k below zero and positive for values above with the magnitude being the largest for k just slightly below or above zero for short expiries. Thus if the approximation (6.44) - (6.50) holds with $\tilde{C}(t)$ as in (6.53) we should by choosing an appropriate level of moneyness be able to control both the sign and the magnitude of the spot/volatility PnL.

Finally let us investigate numerically how well the approximation (6.44) - (6.50) with $\tilde{C}(t)$ as in (6.53) holds in the rough Bergomi model. We conduct the experiment as follows: Fix $k = 0.1$ or $k = -0.1$ and assume that $H^a = 0.15$ but that the market uses an incorrect Hurst exponent $H^i = 0.10$. In figure 6.4 we then plot across the expiries the difference $R_T^r - R_T^i$ (top-left) and for each value of k also $\int_0^T \frac{\partial \delta P}{\partial S \partial \xi(u)} du$ (top-right) and finally the actual and approximate PnL's from equations (6.44) and (6.50) (bottom left and right). If we first consider the case of $k = 0.1$ we as expected find $\int_0^T \frac{\partial \delta P}{\partial S \partial \xi(u)} du$ to be positive. As we have chosen $\rho = -0.9 < 0$ such that the at-the-money skew is negative we should expect the sign of the PnL to be the opposite of the sign of $R_T^r - R_T^i$. Similarly in the case of $k = -0.1$ the factor $\int_0^T \frac{\partial \delta P}{\partial S \partial \xi(u)} du$ is negative and so we should expect the PnL and $R_T^r - R_T^i$ to have the same sign. We confirm this result by comparing the actual PnL's in the bottom figures with the SSR difference in the top-left figure. We also see that the magnitude of the PnL's are heavily influenced by the size of the factor $\int_0^T \frac{\partial \delta P}{\partial S \partial \xi(u)} du$ (top-right). Finally, and importantly, we see that the approximate expression (6.50) seems to work somewhat well in explaining the actual PnL.

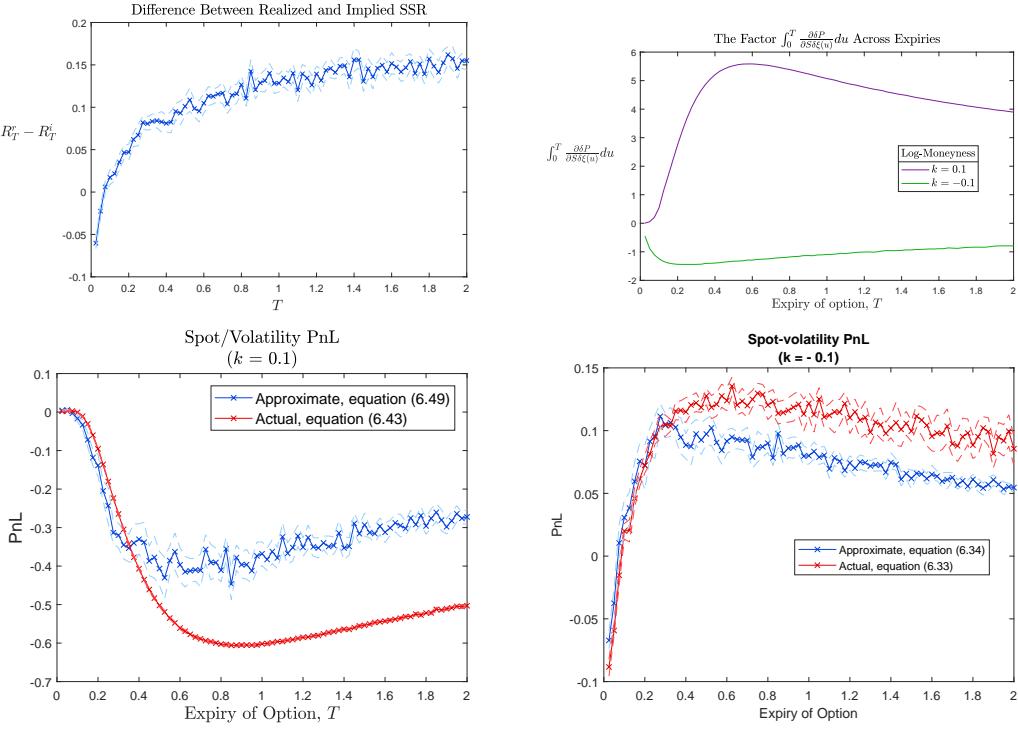


Figure 6.4: Numerical test comparing the spot/volatility PnL from (6.44) with the approximative expression from (6.50) with $\tilde{C}(t) = \frac{1}{T-t} \int_t^T \frac{\partial \delta P}{\partial S \xi(u)} du$. Fixed parameters are: $H^a = 0.15, H^i = 0.10, \eta = 1.9, \rho = -0.9, S_0 = 100, r = 0.05, q = 0.02$ and $\xi_0(t) = 0.04$ for all t . Lines shown are the average of 25 repeated estimates each computed using 10.000 Monte Carlo paths (half of which are antithetic). Dashed lines are 95 % confidence intervals computed by bootstrapping with 100.000 samples.

We can conclude a couple of things: First of all it turns out to be more natural to trade the difference between realized and implied SSR rather than the difference in Hurst exponents. We have also argued and demonstrated that this difference is realized via the spot/volatility PnL and while we have only checked the validity of the approximation numerically for the rough Bergomi model the theoretical argument should hold for any continuous model on forward variance curve form. Finally, we have argued that one should be careful picking the correct strike for the hedged option to achieve the correct sign of this PnL and to influence the magnitude of it. As we have not yet properly handled the volatility-gamma PnL we think there is plenty of room for further research into how one can properly trade the SSR under the rough Bergomi model.

Chapter 7

Conclusion

In this thesis we have considered the new class of rough volatility models with a special focus on the rough Bergomi model. We started by investigating the volatility process empirically via high frequency stock price data on the SPY. We found that log-volatility is a very rough process that is also unifractal with a Hurst exponent of $H \approx 0.1$. We also found evidence of the log-increments of volatility to be almost Gaussian when sampled once per day. Though we found spurious evidence of long-memory we have been unable to finally settle this question. Disregarding the long memory property and settling for the Gaussian approximation we proposed to model volatility via a fractional Brownian motion with a Hurst exponent of less than $\frac{1}{2}$.

We then used a deterministic change of measure to arrive at the rough Bergomi model. While we find that the non-Markovian nature of the fractional Brownian motion can be captured in the forward variance curve which can be estimated from option prices we don't find any feasible markovian methods of computing option prices within the model. Thus we find the best one can do is to rely on Monte Carlo simulation to estimate option prices. In that regard we investigated various simulation techniques and found that while an exact simulation scheme for simulating the volatility process is available it is also slow and thus impractical for large scale Monte Carlo computations. This problem can be solved by using instead a hybrid scheme which offers both speed and sufficient distributional accuracy. With this in mind and using the various variance reduction techniques proposed we think price estimation under the model is now also feasible in a real world setting.

We also considered calibration. Estimating the forward variance curve from option prices we find that one is left with a three-dimensional optimization problem. Unfortunately even this problem is infeasible in practise. Using among other things an accurate short expiry skew expansion from (Fukasawa 2015) we were nonetheless able to construct a fast calibration method. We had some success with our method although a little more research is needed to obtain reasonable parameter estimates in light of the options data that is available for the S&P 500. Regardless of this minor problem we conclude that calibration now is only a question of minutes rather than hours. In terms of the fit we found that the rough Bergomi model is able to capture the power-law decay of skew observed in practise.

Finally we looked into hedging in the model. We find that this also is made somewhat more complicated as the model is only Markov once conditioned on the entire forward variance curve. Here we also considered the dynamic properties of the model. We find that the rough Bergomi model generally displays a SSR of $\frac{3}{2} + H$ and thus is able to capture the empirically observed values of SSR that are around 1.5. Finally we presented some thoughts on how the difference between realized and implied SSR can be realized as a PnL. We find that this to be less tractable due to the non-markovian nature of the model. Up to some approximations we were however able to isolate this as PnL through the spot-volatility term and we were able to successfully confirm this result numerically.

Appendix A

Numerical Results

A.1 Seasonality Factor by Year

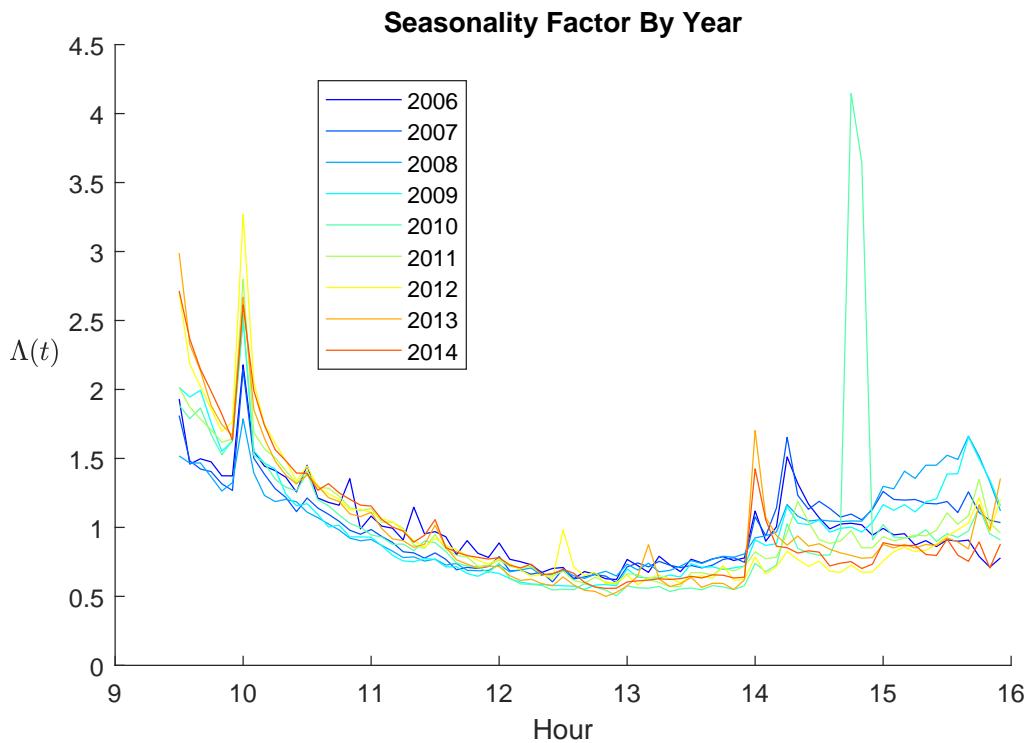


Figure A.1: Intraday seasonality factor calculated by taking averages across each year in question. Underlying data is 5 minute sampled realized variance estimates. The spike in the factors for 2010 is due to the Flash Crash on May 6th 2010 which occurred around 14:30 in the afternoon.

A.2 Statistics: Volatility by Sampling Frequency

In table A.1 we show some statistics of the estimated volatility process when you vary the sampling frequency.

Sampling frequency (Δ)	Number of observations	Mean of realized variance	Skew of log-volatility increments	Kurtosis of log-volatility increments
5 min.	174,642	0.0222	0.1596	4.4493
10 min.	87,399	0.0228	0.2677	4.9856
15 min.	58,266	0.0232	0.3487	5.5365
30 min.	29,146	0.0240	0.6007	6.4839
65 min.	13,458	0.0249	0.8049	6.6893
130 min.	6,729	0.0255	0.6847	5.7018
1 day	2,243	0.0262	0.1810	4.2066

Table A.1: Sample statistics on realized variance and log volatility increments using different sampling frequencies. Numbers are annualized using 252 trading days per year. The skew and kurtosis columns are computed using the deseasonalised volatility estimates.

A brief comment on the third column: As we can see the realized variance estimates tends to decrease, although very little, when the sampling frequency increases. This is initially counterintuitive as we should, despite the pre-averaging, expect microstructure noise to dominate at the very low frequencies thus giving higher integrated variances. However, since our estimator already includes a bias-correction term this will most likely cancel some, all or too much of this effect. We think the latter is why the estimates decrease slightly.

A.3 Histograms of Log-Volatility

Recall the scaling relationship

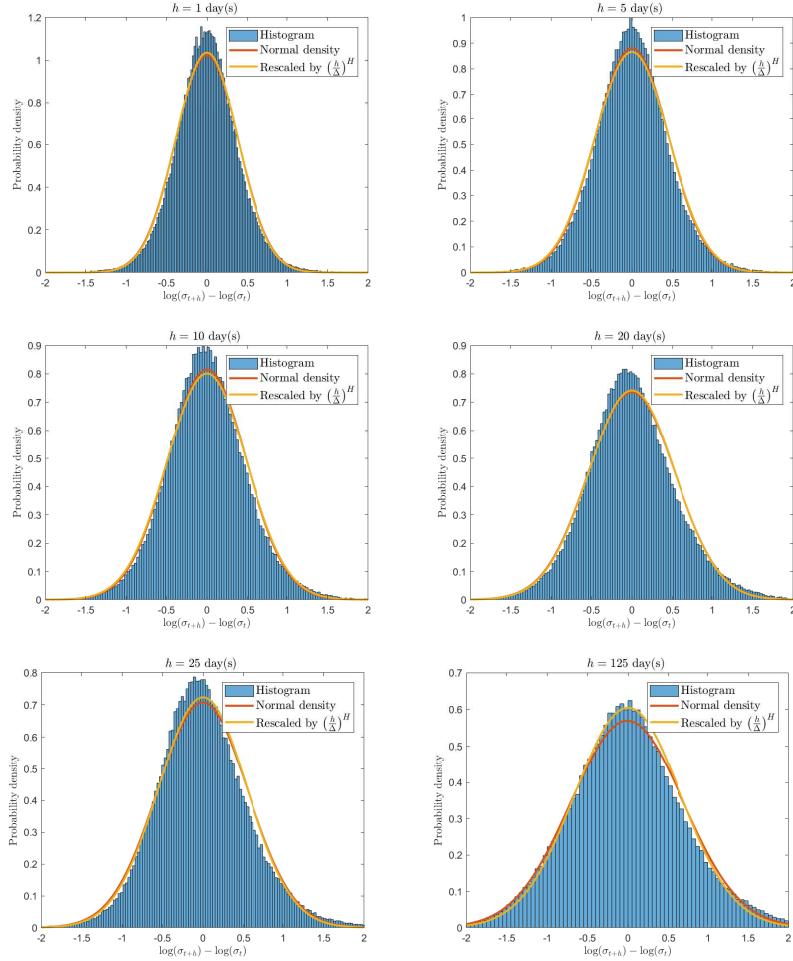
$$E(|\log(\sigma_{t+h}) - \log(\sigma_t)|^q) = K_q h^{qH} \quad (\text{A.1})$$

for some increment length h that is a multiple of the sampling frequency Δ . As can be seen below the $(\frac{h}{\Delta})^H$ scaled process of log-volatility increments ($\log(\sigma_{t+\Delta}) - \log(\sigma_t)$) will adhere to the same relationship as $\log(\sigma_{t+h}) - \log(\sigma_t)$:

$$E\left(\left|\left(\frac{h}{\Delta}\right)^H (\log(\sigma_{t+\Delta}) - \log(\sigma_t))\right|^q\right) = \left(\frac{h}{\Delta}\right)^{qH} K_q \Delta^{qH} = K_q h^{qH}. \quad (\text{A.2})$$

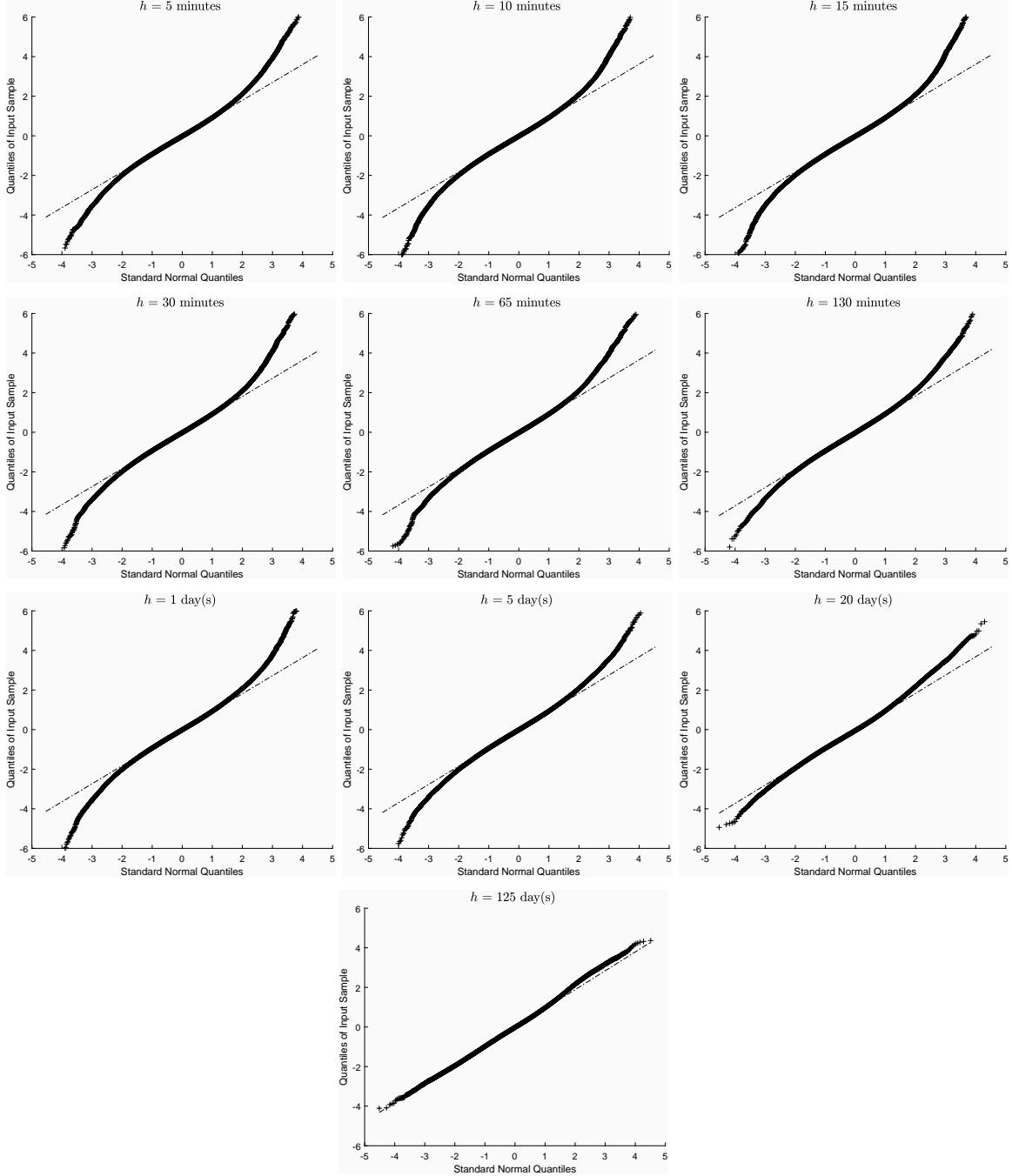
As this holds for any q we should expect the density of the increments $\log(\sigma_{t+h}) - \log(\sigma_t)$ to be the same as the density of $(\frac{h}{\Delta})^H (\log(\sigma_{t+\Delta}) - \log(\sigma_t))$.

This relationship is exactly what we consider in the below plots. Here the underlying data is the $\Delta = 5$ minutes sampled volatility estimates and h is then a multiple of 5 minutes. The red curves are densities of normal distributions fitted to the samples $\log(\sigma_{t+h}) - \log(\sigma_t)$ and the yellow curves are densities of normal distributions fitted to the rescaled samples $(\frac{h}{\Delta})^H (\log(\sigma_{t+\Delta}) - \log(\sigma_t))$.



A.4 Log-Volatility Increments

Below we show QQ plots of the log volatility increments $\log(\sigma_{t+h}) - \log(\sigma_t)$ for various choices of h . To allow for a better comparison we have normalized the samples so they all have mean 0 and standard deviation 1. The sampling frequency is $\Delta = 5$ minutes.



A.5 Roughness and Persistence

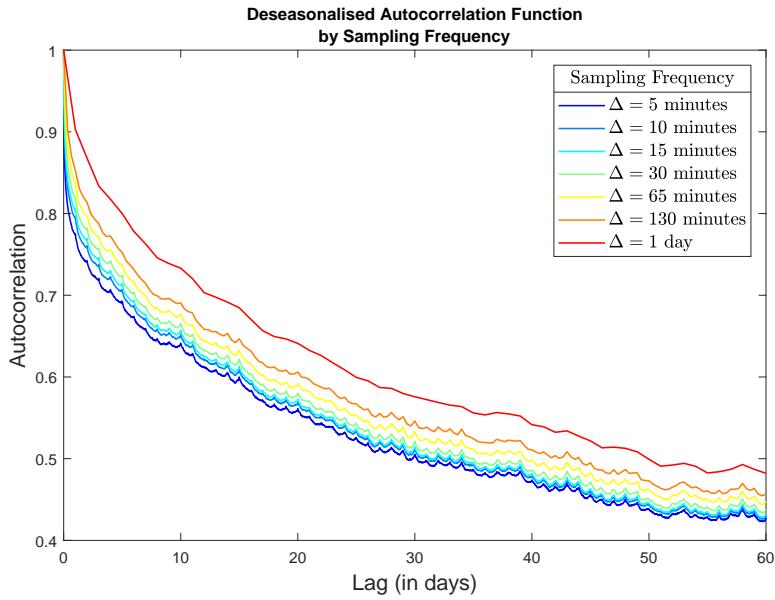


Figure A.2: The autocorrelation function of log-volatility sampled at various frequencies and that after adjusting for seasonality with weights recomputed each month.

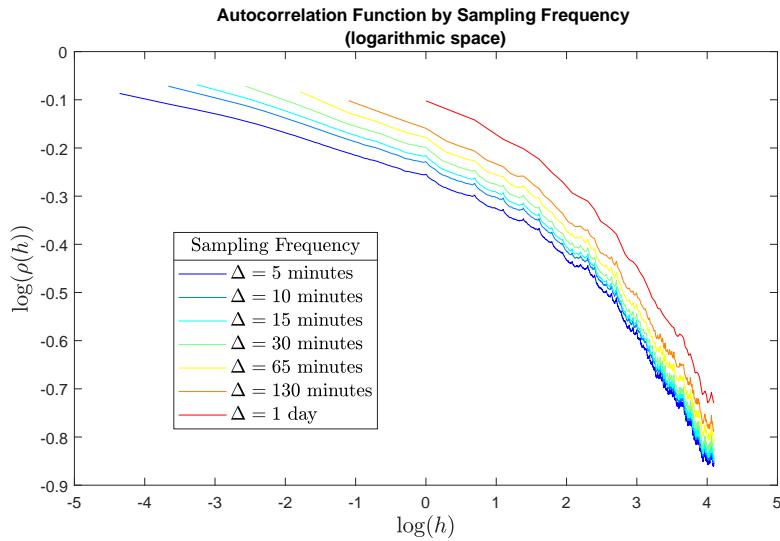


Figure A.3: Plot shows the same as figure A.2 except we have taken log on both axes. Note that $h = 1$ corresponds to 1 day.

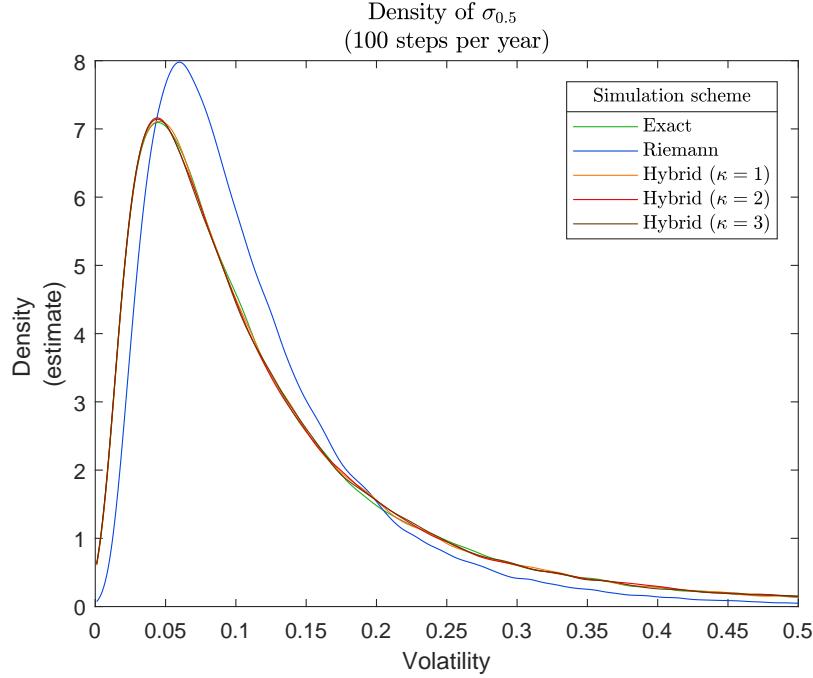
Δ	$\hat{\alpha}$	$\hat{\beta}$
5 min.	-0.3674	0.2495
10 min.	-0.3164	0.2547
15 min.	-0.2924	0.2553
30 min.	-0.2794	0.2542
65 min.	-0.3024	0.2531
130 min.	-0.3151	0.2553
1 day	-0.2772	0.2621

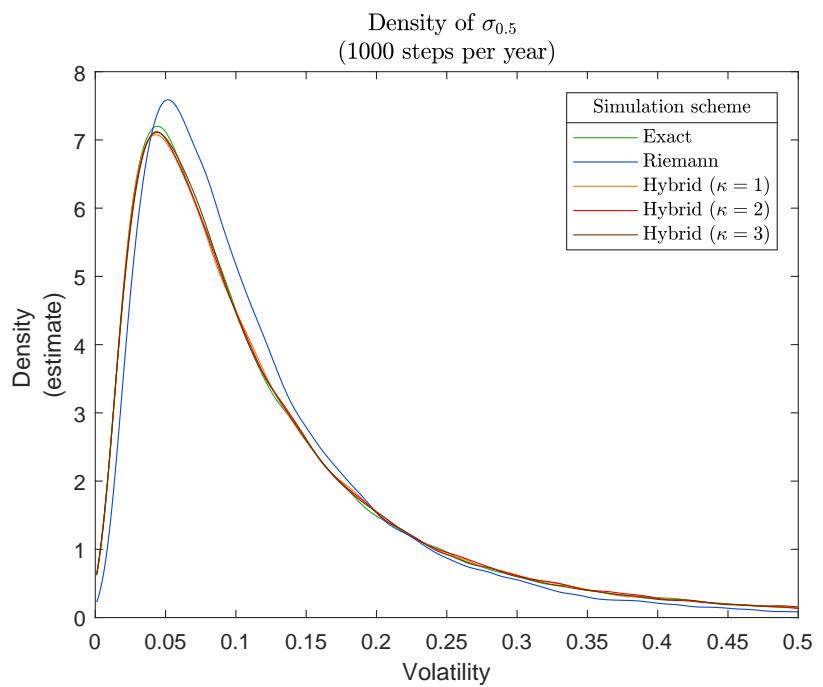
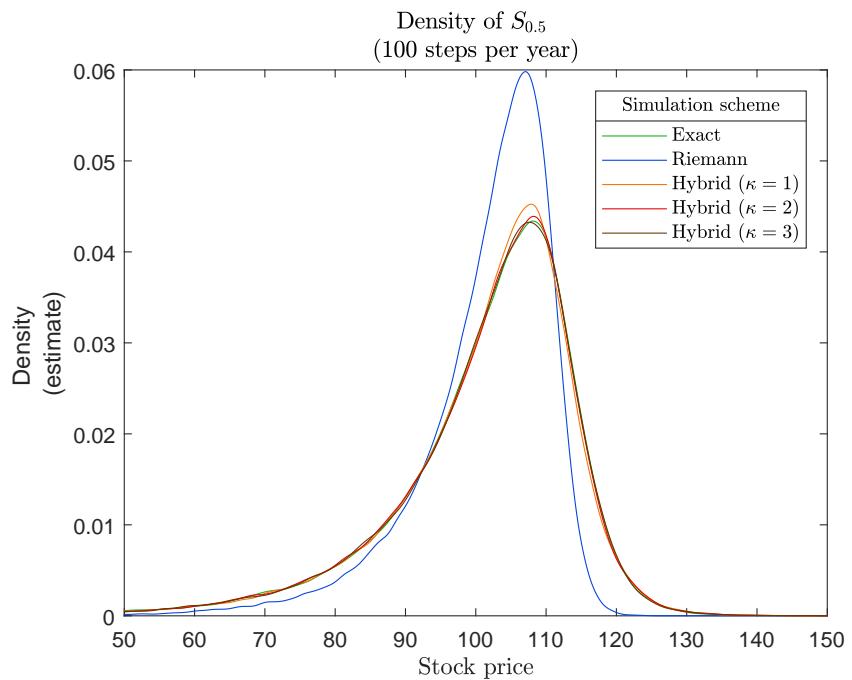
Table A.2: Estimations of the fractal index α and the persistence parameter β using the methods from sections 2.2.3 and 2.2.4 and that across the different sampling frequencies. In particular we use $m = 6$ to estimate α and choose M, M' s.t. $M\Delta = 7$ days and $M'\Delta = 60$ days to estimate β .

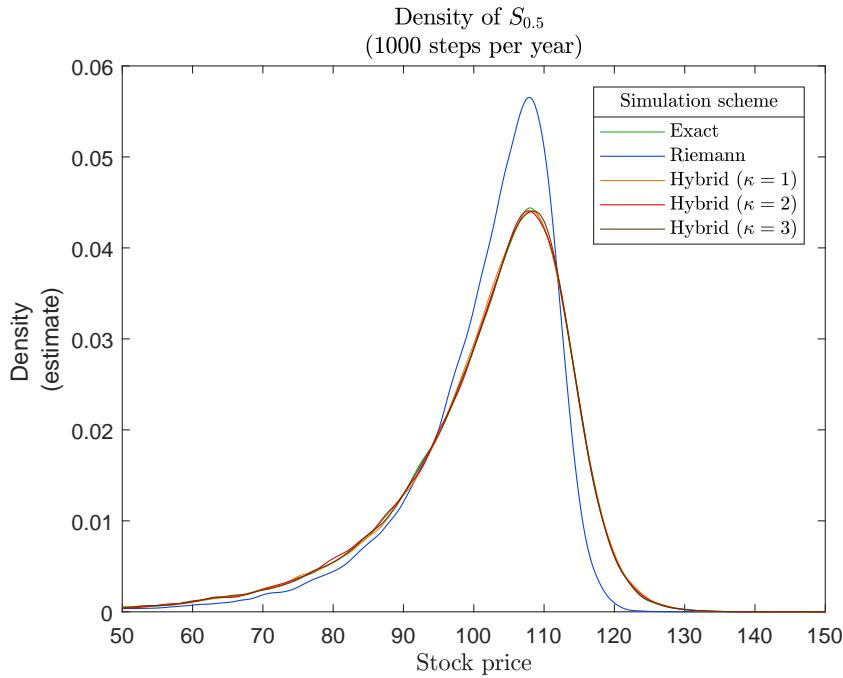
A.6 Bias of Simulation Schemes

Here we present results illustrating the bias of the various simulation schemes we considered. The fixed parameters are $H = 0.1$, $\rho = -0.9$, $\eta = 1.9$, $S_0 = 100$, $r = 0.05$, $q = 0.02$ and $\xi_0(t) = 0.04$ for all t .

Below we show density estimates of the stock price S_t and volatility process σ_t at $t = \frac{1}{2}$ using the various simulation schemes. Estimates are based on Matlab's `ksdensity` function and are all based on 100,000 samples.







We also considered comparing the distributional accuracy of the various schemes by using a two-sided Kolmogorov-Smirnov (KS) test. In particular we sampled 100,000 paths using the exact simulation scheme and that with 1,000 steps per year. Via the two-sided KS test we then compare this sample with samples of size 1,000 from the other simulation schemes and that across different choices of steps per year. Repeating the experiment 1,000 times gives us the average p values in the below tables. If we use a significance level of 5 % any p values below indicates that the KS test rejects the hypothesis of the sample from the scheme as coming from the true distribution as approximated by the exact scheme. Values in parenthesis are standard deviations computed using the central limit theorem. Parameters are as before and we keep $t = \frac{1}{2}$.

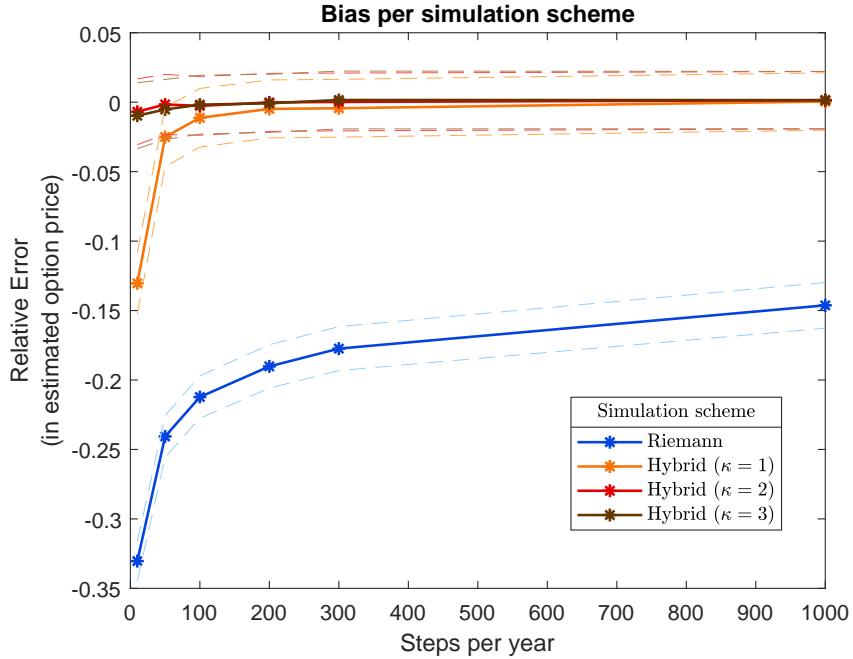
Table A.3: Two-sided Kolmogorov-Smirnov test for samples of $\sigma_{0.5}$

Steps per year	Riemann	Hybrid ($\kappa = 1$)	Hybrid ($\kappa = 2$)	Hybrid ($\kappa = 3$)
10	$2.6 \cdot 10^{-12}$ ($6.9 \cdot 10^{-20}$)	0.5194 ($3.3 \cdot 10^{-7}$)	0.5106 (0.0031)	0.4998 (0.0031)
30	$2.2 \cdot 10^{-7}$ ($1.0 \cdot 10^{-15}$)	0.5114 (0.0027)	0.5070 (0.0085)	0.4806 (0.0081)
50	$7.1 \cdot 10^{-6}$ ($7.6 \cdot 10^{-13}$)	0.5062 (0.0067)	0.4971 (0.0090)	0.5052 (0.0091)
100	$1.7 \cdot 10^{-4}$ ($2.0 \cdot 10^{-10}$)	0.5023 (0.0089)	0.4912 (0.0093)	0.5012 (0.0094)
200	0.0016 ($1.7 \cdot 10^{-9}$)	0.5069 (0.0091)	0.5036 (0.0090)	0.5185 (0.0090)
500	0.0100 ($1.0 \cdot 10^{-7}$)	0.4860 (0.0092)	0.4935 (0.0091)	0.5088 (0.0089)
1000	0.0300 ($9.5 \cdot 10^{-7}$)	0.5091 (0.0091)	0.4904 (0.0092)	0.5140 (0.0093)

Table A.4: Two-sided Kolmogorov-Smirnov test for samples of $S_{0.5}$

Steps per year	Riemann	Hybrid ($\kappa = 1$)	Hybrid ($\kappa = 2$)	Hybrid ($\kappa = 3$)
10	$8.0 \cdot 10^{-20}$ ($6.9 \cdot 10^{-20}$)	$4.2 \cdot 10^{-7}$ ($3.3 \cdot 10^{-7}$)	0.0591 (0.0031)	0.0559 (0.0031)
30	$2.3 \cdot 10^{-14}$ ($1.0 \cdot 10^{-15}$)	0.0430 (0.0027)	0.3020 (0.0085)	0.2742 (0.0081)
50	$1.7 \cdot 10^{-12}$ ($7.6 \cdot 10^{-13}$)	0.1856 (0.0067)	0.3750 (0.0090)	0.3919 (0.0091)
100	$3.3 \cdot 10^{-10}$ ($2.1 \cdot 10^{-10}$)	0.3779 (0.0089)	0.4688 (0.0093)	0.4573 (0.0094)
200	$5.9 \cdot 10^{-9}$ ($1.7 \cdot 10^{-9}$)	0.4612 (0.0091)	0.4901 (0.0090)	0.4789 (0.0090)
500	$3.6 \cdot 10^{-7}$ ($1.0 \cdot 10^{-7}$)	0.5146 (0.0092)	0.4882 (0.0091)	0.5191 (0.0089)
1000	$5.2 \cdot 10^{-6}$ ($9.5 \cdot 10^{-7}$)	0.5097 (0.0091)	0.5014 (0.0092)	0.4984 (0.0093)

We also considered the effect of using the various schemes on estimating option prices via Monte Carlo. Thus keeping the parameters as before and considering an expiry $T = \frac{1}{2}$ option with log-moneyness $k = \log\left(\frac{K}{F_T}\right) = 0$ we estimate option prices via Monte Carlo for each scheme using 1 million paths and that across different numbers of steps per year. We estimate the true option price by using the exact simulation scheme with 1 million paths and 1,000 steps per year. Below we plot the relative error defined as $\frac{\text{Price Estimate} - \text{Actual Price}}{\text{Actual Price}}$ across different numbers of steps per year. Dashed lines are 95 % confidence intervals computed using the central limit theorem.



A.7 The Mixed Estimator: Bias from Choice of V

Here we test whether or not the choice of $V = \sup \left\{ \left(\int_0^T v_t dt \right)_{i=1}^N \right\}$ from subsection 4.4 leads to any bias in the mixed estimator. In the below tables we have estimated both put and call options under the rough Bergomi model using 1) the usual, or base, Monte Carlo estimator (including antithetic paths but otherwise without any variance reduction techniques) and 2) the mixed estimator with V as stated before. The experiment is conducted as follows: We use the hybrid scheme with 200 steps per year to simulate the model. We then estimate each price 100 times using in each case 10,000 paths, half of which are antithetic. The prices shown are averages of those 100 samples. The standard errors (s.e.) are likewise computed from those 100 samples. Using the Black-Scholes delta, we write Δ_{BS} , as a measure of 'moneyness' we have conducted our test across different values of Δ_{BS} . The fixed model parameters are: $H = 0.1$, $\rho = -0.9$, $\eta = 1.9$, $S_0 = 100$, $r = 0.05$, $q = 0.02$ and $\xi_0(t) = 0.04$ for all t .

Table A.5: Call options.

T	Δ_{BS}	k	Mixed (s.e.)	Base (s.e.)	Diff. (s.e.)
0.25	0.1	0.09	0.302 (0.001)	0.301 (0.001)	0.000 (0.001)
0.25	0.5	0.00	3.261 (0.003)	3.255 (0.004)	0.007 (0.002)
0.25	0.9	-0.17	16.492 (0.006)	16.486 (0.008)	0.006 (0.004)
0.50	0.1	0.12	0.400 (0.001)	0.399 (0.002)	0.001 (0.001)
0.50	0.5	0.01	4.420 (0.004)	4.413 (0.005)	0.006 (0.004)
0.50	0.9	-0.26	23.398 (0.009)	23.393 (0.010)	0.006 (0.006)
1.00	0.1	0.16	0.547 (0.001)	0.546 (0.003)	0.001 (0.002)
1.00	0.5	0.01	5.965 (0.005)	5.960 (0.007)	0.005 (0.005)
1.00	0.9	-0.41	34.543 (0.011)	34.537 (0.013)	0.007 (0.009)
2.00	0.1	0.22	0.729 (0.002)	0.730 (0.003)	-0.001 (0.003)
2.00	0.5	0.01	8.043 (0.007)	8.043 (0.010)	0.000 (0.007)
2.00	0.9	-0.85	57.573 (0.018)	57.564 (0.021)	0.009 (0.012)

Table A.6: Put options.

T	Δ_{BS}	k	Mixed (s.e.)	Base (s.e.)	Diff. (s.e.)
0.25	-0.1	-0.16	0.589 (0.003)	0.588 (0.003)	0.001 (0.001)
0.25	-0.5	0.00	3.621 (0.005)	3.618 (0.006)	0.003 (0.003)
0.25	-0.9	0.09	9.412 (0.006)	9.414 (0.007)	-0.002 (0.004)
0.50	-0.1	-0.24	0.787 (0.003)	0.787 (0.004)	0.001 (0.002)
0.50	-0.5	0.01	5.088 (0.008)	5.087 (0.009)	0.001 (0.004)
0.50	-0.9	0.12	13.244 (0.009)	13.247 (0.011)	-0.003 (0.006)
1.00	-0.1	-0.33	1.050 (0.005)	1.049 (0.005)	0.001 (0.002)
1.00	-0.5	0.02	7.219 (0.011)	7.216 (0.012)	0.004 (0.005)
1.00	-0.9	0.17	19.218 (0.013)	19.217 (0.014)	0.001 (0.008)
2.00	-0.1	-0.48	1.256 (0.005)	1.255 (0.006)	0.001 (0.002)
2.00	-0.5	0.03	10.382 (0.015)	10.386 (0.017)	-0.005 (0.007)
2.00	-0.9	0.26	29.631 (0.018)	29.634 (0.022)	-0.004 (0.010)

A.8 Implied Yields

Expiry	# Put/Call pairs	Yield	# Out-of-the-money options
0.02	49	-1.85 %	53
0.04	66	-2.10 %	88
0.10	94	-0.44 %	149
0.18	102	-0.45 %	155
0.25	103	-0.16 %	138
0.29	30	-0.05 %	34
0.50	35	0.21 %	44
0.75	37	0.40 %	48
1.27	37	0.46 %	51
1.77	28	0.53 %	37
2.27	39	0.59 %	51
Total	620		848

Dividend yield	2.39 %
-----------------------	---------------

A.9 SVI

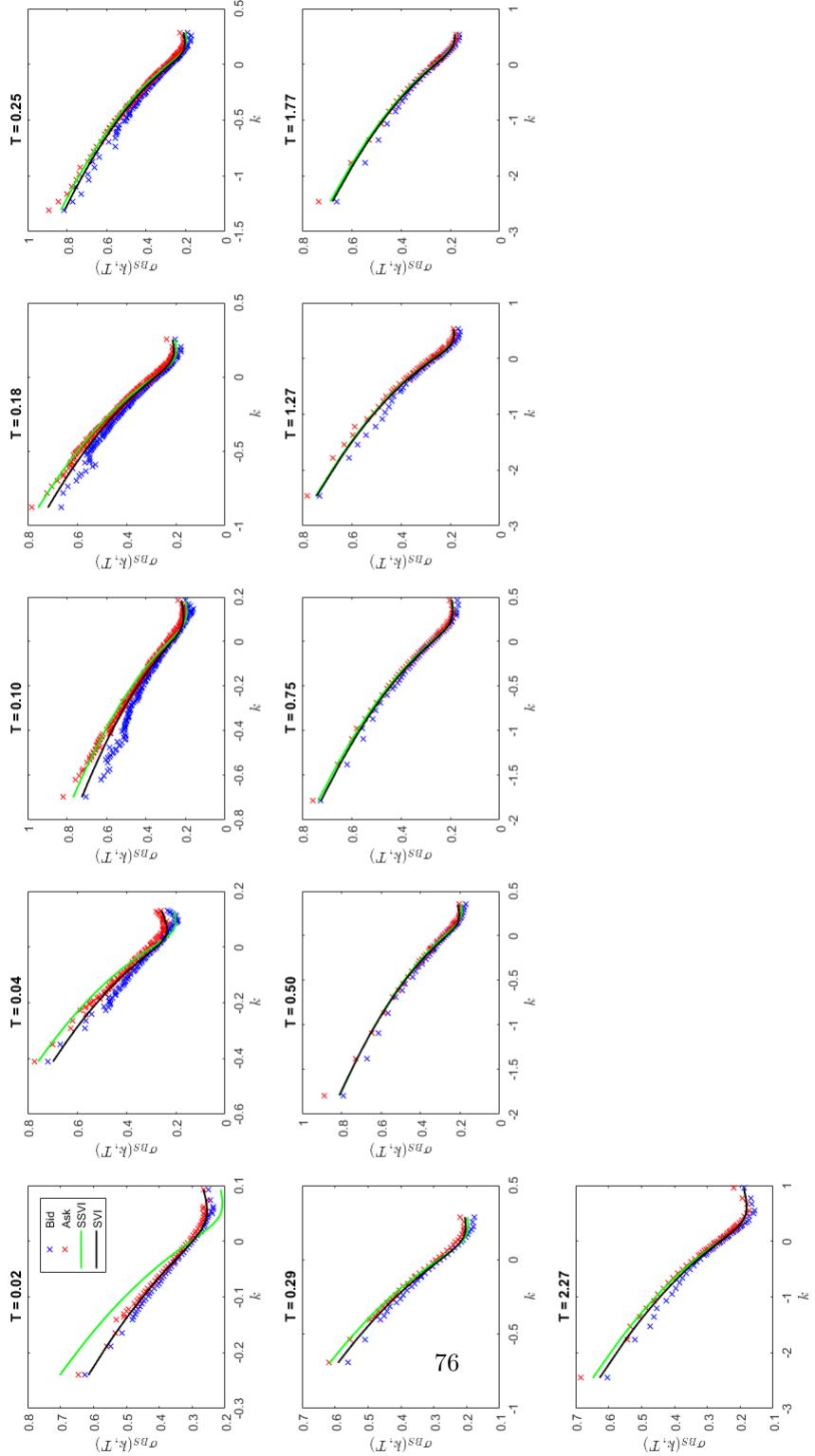


Figure A.4: Full plot of the SSVI and SVI fits as explained in chapter 5.

Table A.7: Natural parameters of full SVI fit

T	Δ	μ	ρ	ω	ζ
0.02	0.0000	0.0000	-0.5216	0.0020	18.8303
0.04	0.0000	0.0000	-0.5360	0.0032	17.3407
0.10	0.0000	0.0000	-0.6283	0.0074	11.2694
0.18	0.0000	0.0000	-0.6817	0.0143	7.5134
0.25	0.0000	0.0000	-0.6765	0.0193	7.1571
0.29	0.0000	0.0000	-0.6884	0.0228	6.0908
0.50	0.0000	0.0000	-0.6502	0.0362	5.6363
0.75	0.0000	0.0000	-0.7119	0.0554	4.0838
1.27	0.0000	0.0000	-0.7175	0.0893	3.2799
1.77	0.0000	0.0000	-0.7149	0.1194	2.8056
2.27	0.0000	0.0000	-0.7205	0.1539	2.3297

Table A.8: Raw parameters of full SVI fit

T	a	b	m	ρ	σ
0.02	0.0007	0.0184	0.0277	-0.5216	0.0453
0.04	0.0011	0.0276	0.0309	-0.5360	0.0487
0.10	0.0022	0.0415	0.0557	-0.6283	0.0690
0.18	0.0038	0.0538	0.0907	-0.6817	0.0974
0.25	0.0052	0.0692	0.0945	-0.6765	0.1029
0.29	0.0060	0.0695	0.1129	-0.6900	0.1185
0.50	0.0105	0.1021	0.1154	-0.6502	0.1348
0.75	0.0137	0.1132	0.1743	-0.7119	0.1720
1.27	0.0217	0.1464	0.2187	-0.7175	0.2124
1.77	0.0292	0.1675	0.2548	-0.7149	0.2492
2.27	0.0370	0.1793	0.3093	-0.7205	0.2976

Table A.9: Jumpwing parameters of full SVI fit

T	v_T	ψ_T	p_T	c_T	\tilde{v}_T
0.02	0.0893	-0.2172	0.6337	0.1993	0.0650
0.04	0.0775	-0.2622	0.7514	0.2270	0.0552
0.10	0.0727	-0.3037	0.7872	0.1797	0.0440
0.18	0.0804	-0.3063	0.7558	0.1431	0.0431
0.25	0.0759	-0.3366	0.8341	0.1610	0.0412
0.29	0.0787	-0.3168	0.7771	0.1434	0.0414
0.50	0.0719	-0.3488	0.8852	0.1877	0.0415
0.75	0.0736	-0.3423	0.8231	0.1385	0.0363
1.27	0.0703	-0.3515	0.8415	0.1384	0.0341
1.77	0.0675	-0.3466	0.8314	0.1382	0.0330
2.27	0.0679	-0.3293	0.7862	0.1277	0.0326

A.10 Rough Bergomi Model: Effect of Parameters on The Smile

Below we illustrate the effect of the three parameters (H, η, ρ) on the implied volatility smiles produced by the rough Bergomi model:

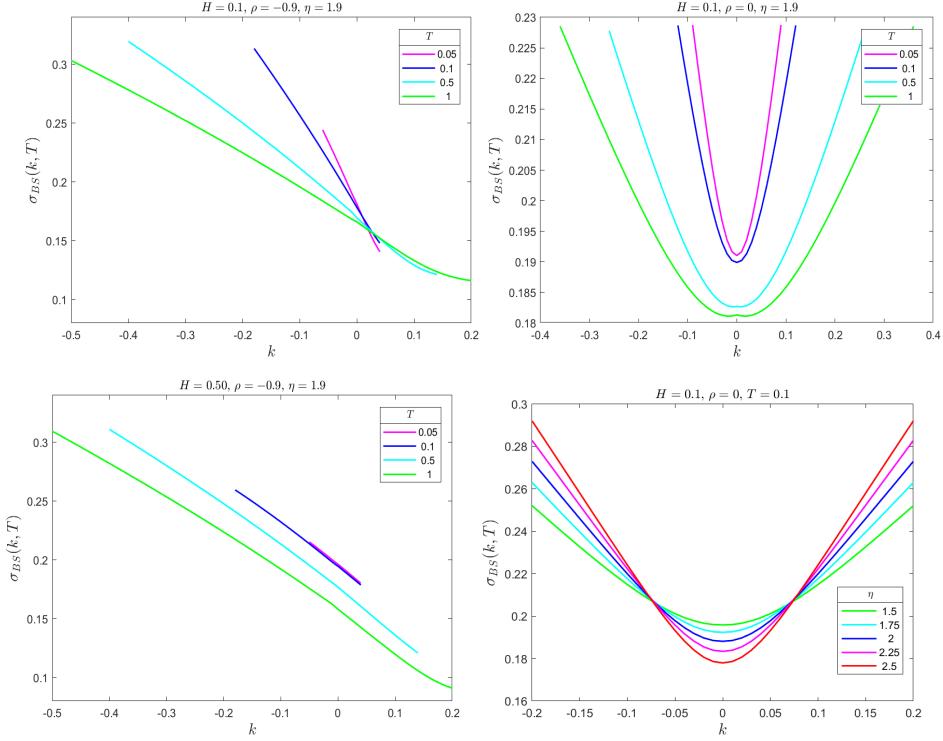


Figure A.5: Effect on the implied volatility smiles of varying the three main parameters H, ρ and η . The following quantities are fixed: $S_0 = 100, r = 0.05, q = 0.02$ and $\xi_0(t) = 0.04$ for all t . Smiles are estimated using the mixed estimator with 400.000 paths (200.000 of which are antithetic).

A.11 Extra Plots of Rough Bergomi Fit

Here we show some extra plots illustrating the fitted rough Bergomi model from chapter 5.

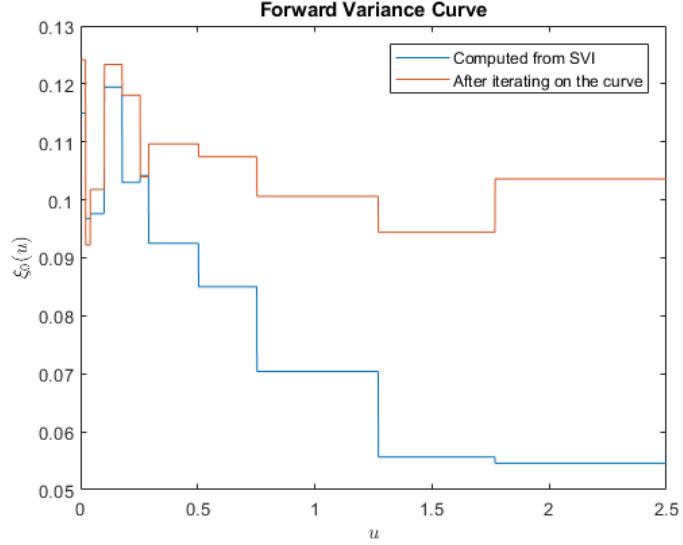


Figure A.6: Forward variance curve before and after having iterated on the curve to match the empirical at-the-money implied volatilities.

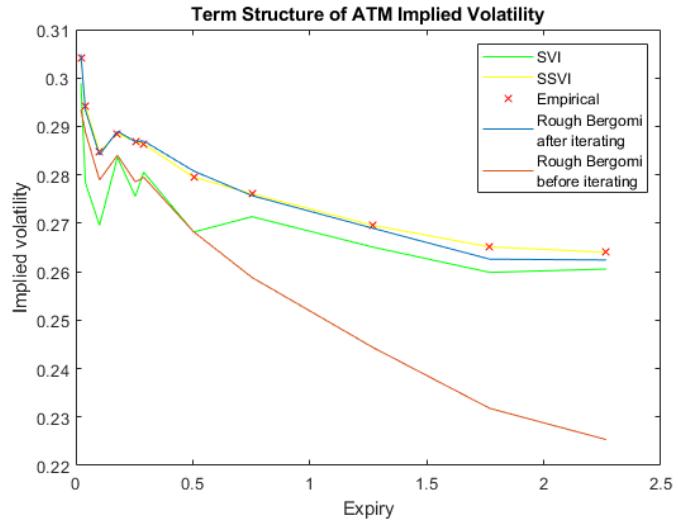


Figure A.7: Empirical at-the-money implied volatilities as well as those computed from the SSVI and SVI fits and also those obtained from the rough Bergomi fit both before and after having iterated on the forward variance curve to match the empirical ones.

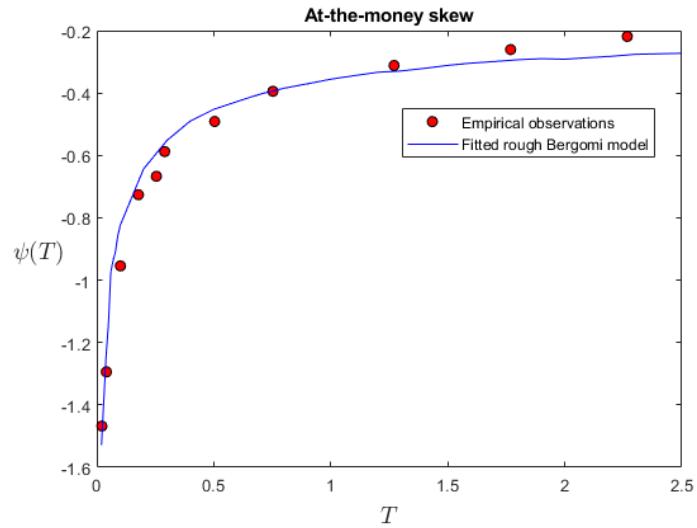


Figure A.8: Empirically observed at-the-money skew as extracted from the SVI fit as well as skew from the fitted Rough Bergomi Model. The latter is estimated using the mixed Monte Carlo estimator with 100.000 paths (50.000 of which are antithetic).

Appendix B

Theory

B.1 Exact Simulation: Covariance Matrix

Here we prove the closed form expressions for the covariance matrix from section 4.2. Note first that for $v \geq u$:

$$\text{Cov}(W_v, W_u) = \text{Cov}(W_v - W_u + W_u, W_u) \quad (\text{B.1})$$

$$= \text{Cov}(W_v - W_u, W_u) + \text{Cov}(W_u, W_u) \quad (\text{B.2})$$

$$= 0 + \text{Var}(W_u) \quad (\text{B.3})$$

$$= u. \quad (\text{B.4})$$

Thus generally $\text{Cov}(W_v, W_u) = \min(u, v)$.

We furthermore get

$$E(\tilde{W}_u \tilde{W}_u) = E\left(\sqrt{2H} \int_0^u \frac{1}{(u-s)^\gamma} dW_s \cdot \sqrt{2H} \int_0^u \frac{1}{(u-s)^\gamma} dW_s\right) \quad (\text{B.5})$$

$$= 2H \int_0^u \frac{1}{(u-s)^{2\gamma}} ds \quad (\text{B.6})$$

$$= 2H \int_u^0 -l^{-2\gamma} dl \quad (\text{B.7})$$

$$= 2H \int_0^u l^{-2\gamma} dl \quad (\text{B.8})$$

$$= 2H \left[\frac{1}{-2\gamma+1} l^{-2\gamma+1} \right]_0^u \quad (\text{B.9})$$

$$= \frac{2H}{-2\gamma+1} u^{-2\gamma+1} \quad (\text{B.10})$$

$$= u^{2H}. \quad (\text{B.11})$$

In the above we changed variables to $l = u - s$ and used that $\gamma = \frac{1}{2} - H$ by definition.

Assume now again $v \geq u$. Then we get the following:

$$E(\tilde{W}_u \tilde{W}_v) = E\left(\sqrt{2H} \int_0^u \frac{1}{(u-s)^\gamma} dW_s \cdot \sqrt{2H} \int_0^v \frac{1}{(v-s)^\gamma} dW_s\right) \quad (\text{B.12})$$

$$= 2H \cdot E\left(\int_0^u \frac{1}{(u-s)^\gamma} dW_s \cdot \int_0^v \frac{1}{(v-s)^\gamma} dW_s\right) \quad (\text{B.13})$$

$$= 2H \cdot E\left(\int_0^u \frac{1}{(u-s)^\gamma} dW_s \cdot \left(\int_u^v \frac{1}{(v-s)^\gamma} dW_s + \int_0^u \frac{1}{(v-s)^\gamma} dW_s\right)\right) \quad (\text{B.14})$$

$$= 2H \cdot E\left(\underbrace{\int_0^u \frac{1}{(u-s)^\gamma} dW_s \cdot \int_u^v \frac{1}{(v-s)^\gamma} dW_s}_{=0}\right) \quad (\text{B.15})$$

$$+ 2H \cdot E\left(\int_0^u \frac{1}{(u-s)^\gamma} dW_s \cdot \int_0^u \frac{1}{(v-s)^\gamma} dW_s\right) \quad (\text{B.16})$$

$$= 2H \cdot \int_0^u \frac{1}{(u-s)^\gamma (v-s)^\gamma} ds. \quad (\text{B.17})$$

The expression in the fourth equality is zero as the two stochastic integrals are independent and have mean zero. The last equality is easily show by an application of Ito on the product of the two stochastic integrals.

Now consider the change of variable $x = u^{-1}s$ which allows us to write:

$$2H \cdot \int_0^u \frac{1}{(u-s)^\gamma (v-s)^\gamma} ds \quad (\text{B.18})$$

$$= 2H \cdot \int_0^1 \frac{1}{(u-ux)^\gamma (v-ux)^\gamma} u dx \quad (\text{B.19})$$

$$= 2H \cdot u \cdot u^{-2\gamma} \int_0^1 \frac{1}{(1-x)^\gamma (\frac{v}{u}-x)^\gamma} dx \quad (\text{B.20})$$

$$= u^{2H} \cdot 2H \cdot \int_0^1 \frac{1}{(1-x)^\gamma (\frac{v}{u}-x)^\gamma} dx \quad (\text{B.21})$$

$$= u^{2H} \cdot 2H \cdot \left(\frac{u}{v}\right)^\gamma \cdot \int_0^1 \frac{1}{(1-x)^\gamma (1-x\frac{u}{v})^\gamma} dx \quad (\text{B.22})$$

$$= u^{2H} \frac{1-2\gamma}{1-\gamma} \left(\frac{u}{v}\right)^\gamma {}_2F_1\left(1, \gamma, 2-\gamma, \frac{u}{v}\right). \quad (\text{B.23})$$

The last equality follows by recalling Euler's integral representation of the confluent hypergeometric function ${}_2F_1$, see (Olver 1997) section 9.

Let now $v \geq u$. Then we further more get:

$$E(\tilde{W}_v W_u) = E\left(\sqrt{2H} \int_0^v \frac{1}{(v-s)^\gamma} dW_{1,s} \cdot W_{2,u}\right) \quad (\text{B.24})$$

$$= E\left(\sqrt{2H} \int_0^v \frac{1}{(v-s)^\gamma} dW_{1,s} \cdot (\rho W_{1,u} + \sqrt{1-\rho^2} W_{\perp,u})\right) \quad (\text{B.25})$$

$$= \sqrt{2H} \rho E\left(\left(\int_u^v \frac{1}{(v-s)^\gamma} dW_{1,s}\right) W_{1,u}\right) \quad (\text{B.26})$$

$$= \sqrt{2H} \rho \int_0^u \frac{1}{(v-s)^\gamma} ds \quad (\text{B.27})$$

$$= \sqrt{2H} \rho \int_{v-u}^v \frac{1}{l^\gamma} dl \quad (\text{B.28})$$

$$= \sqrt{2H} \rho \frac{1}{-\gamma+1} [v^{-\gamma+1} - (v-u)^{-\gamma+1}] \quad (\text{B.29})$$

$$= \sqrt{2H} \rho \frac{1}{H+\frac{1}{2}} (v^{H+1/2} - (v-u)^{H+1/2}) \quad (\text{B.30})$$

$$= \rho D_H \left\{ v^{H+1/2} - (v-u)^{H+1/2} \right\}. \quad (\text{B.31})$$

Here we have changed variables $l = v - s$ and defined $D_H = \frac{\sqrt{2H}}{H+1/2}$.

In a similar style of fashion we get

$$E(W_v \tilde{W}_u) = E\left(W_{2,v} \cdot \sqrt{2H} \int_0^u \frac{1}{(u-s)^\gamma} dW_{1,s}\right) \quad (\text{B.32})$$

$$= \sqrt{2H} \rho \int_0^u \frac{1}{(u-s)^\gamma} ds \quad (\text{B.33})$$

$$= \sqrt{2H} \rho \frac{1}{-\gamma+1} u^{-\gamma+1} \quad (\text{B.34})$$

$$= \rho D_H u^{H+1/2}. \quad (\text{B.35})$$

These two formulas can be combined as

$$E(\tilde{W}_v W_u) = \rho D_H \left(v^{H+1/2} - (v - \min(u, v))^{H+1/2} \right) \quad (\text{B.36})$$

which then hold for any $u, v \geq 0$.

B.2 Hybrid Scheme: Covariance Matrix

Here we prove the closed form expressions for the covariance matrix from section 4.3.

Let therefore $\kappa \in \{1, 2, 3, \dots\}$.

Without loss of generality it suffices to compute the $(\kappa + 1) \times (\kappa + 1)$ dimensional covariance matrix Σ of

$$\mathbf{W}_0^n = (W_0^n, W_{0,1}^n, \dots, W_{0,\kappa}^n) \quad (\text{B.37})$$

where one may recall that

$$W_0^n = \int_0^{\frac{1}{n}} dW_{1,s} \quad (\text{B.38})$$

and

$$W_{0,j}^n = \int_0^{\frac{1}{n}} \left(\frac{j}{n} - s \right)^{-\gamma} dW_{1,s} \quad (\text{B.39})$$

for $j \in \{1, 2, \dots, \kappa\}$.

First we get

$$\Sigma_{1,1} = \frac{1}{n} \quad (\text{B.40})$$

as $W_{1,\frac{1}{n}} \sim \mathcal{N}(0, \frac{1}{n})$.

Let now $j \in \{2, \dots, \kappa + 1\}$.

As $E(W_{0,j-1}^n) = 0$ and using the Ito isometry we also get

$$\Sigma_{j,j} = E([W_{0,j-1}^n]^2) \quad (\text{B.41})$$

$$= \int_0^{\frac{1}{n}} \left(\frac{j-1}{n} - s \right)^{-2\gamma} ds \quad (\text{B.42})$$

$$= -\frac{1}{1-2\gamma} \left[\left(\frac{j-1}{n} - s \right)^{1-2\gamma} \right]_0^{\frac{1}{n}} \quad (\text{B.43})$$

$$= -\frac{1}{1-2\gamma} \left(\left(\frac{j-2}{n} \right)^{1-2\gamma} - \left(\frac{j-1}{n} \right)^{1-2\gamma} \right) \quad (\text{B.44})$$

$$= \frac{(j-1)^{1-2\gamma} - (j-2)^{1-2\gamma}}{(1-2\gamma) \cdot n^{1-2\gamma}}. \quad (\text{B.45})$$

Consider again some $j \in \{2, \dots, \kappa + 1\}$.

Since $E(W_0^n) = E(W_{0,j-1}^n) = 0$ we get

$$\Sigma_{1,j} = E(W_0^n W_{0,j-1}^n) \quad (\text{B.46})$$

$$= \int_0^{\frac{1}{n}} \left(\frac{j-1}{n} - s \right)^{-\gamma} ds \quad (\text{B.47})$$

$$= -\frac{1}{1-\gamma} \left[\left(\frac{j-1}{n} - s \right)^{1-\gamma} \right]_0^{\frac{1}{n}} \quad (\text{B.48})$$

$$= -\frac{1}{1-\gamma} \left(\left(\frac{j-2}{n} \right)^{1-\gamma} - \left(\frac{j-1}{n} \right)^{1-\gamma} \right) \quad (\text{B.49})$$

$$= \frac{(j-1)^{1-\gamma} - (j-2)^{1-\gamma}}{(1-\gamma) \cdot n^{1-\gamma}} \quad (\text{B.50})$$

where the second equality is easily shown by an application of Ito's lemma.

Let now $j, k \in \{2, \dots, \kappa + 1\}$, $j < k$. Then

$$\Sigma_{j,k} = E(W_{0,j-1} W_{0,k-1}) \quad (\text{B.51})$$

$$= \int_0^{\frac{1}{n}} \left(\frac{j-1}{n} - s \right)^{-\gamma} \left(\frac{k-1}{n} - s \right)^{-\gamma} ds \quad (\text{B.52})$$

$$= \int_0^1 \left(\frac{j-1}{n} - \frac{x}{n} \right)^{-\gamma} \left(\frac{k-1}{n} - \frac{x}{n} \right)^{-\gamma} n^{-1} dx \quad (\text{B.53})$$

$$= n^{2\gamma-1} \int_0^1 (j-1-x)^{-\gamma} (k-1-x)^{-\gamma} dx \quad (\text{B.54})$$

$$= n^{2\gamma-1} \left(\int_0^{j-1} (j-1-x)^{-\gamma} (k-1-x)^{-\gamma} dx + \int_{j-1}^1 (j-1-x)^{-\gamma} (k-1-x)^{-\gamma} dx \right) \quad (\text{B.55})$$

$$= n^{2\gamma-1} \left(\int_0^{j-1} (j-1-x)^{-\gamma} (k-1-x)^{-\gamma} dx - \int_1^{j-1} (j-1-x)^{-\gamma} (k-1-x)^{-\gamma} dx \right) \quad (\text{B.56})$$

$$= n^{2\gamma-1} \left(\int_0^{j-1} (j-1-x)^{-\gamma} (k-1-x)^{-\gamma} dx - \int_0^{j-2} (j-2-y)^{-\gamma} (k-2-y)^{-\gamma} dy \right) \quad (\text{B.57})$$

$$= n^{2\gamma-1} \frac{1}{1-2\gamma} (j-1)^{1-2\gamma} \frac{1-2\gamma}{1-\gamma} \left(\frac{j-1}{k-1} \right)^\gamma {}_2F_1 \left(1, \gamma, 2-\gamma, \frac{j-1}{k-1} \right) \quad (\text{B.58})$$

$$- n^{2\gamma-1} \frac{1}{1-2\gamma} (j-2)^{1-2\gamma} \frac{1-2\gamma}{1-\gamma} \left(\frac{j-2}{k-2} \right)^\gamma {}_2F_1 \left(1, \gamma, 2-\gamma, \frac{j-2}{k-2} \right) \quad (\text{B.59})$$

$$= \frac{n^{2\gamma-1}}{1-\gamma} \left((j-1)^{1-\gamma} (k-1)^{-\gamma} {}_2F_1 \left(1, \gamma, 2-\gamma, \frac{j-1}{k-1} \right) - (j-2)^{1-\gamma} (k-2)^{-\gamma} {}_2F_1 \left(1, \gamma, 2-\gamma, \frac{j-2}{k-2} \right) \right). \quad (\text{B.60})$$

In the above the second equality again follows by an application of Ito's lemma. Furthermore in the third equality we used the change of variable $x = ns$, in the seventh equality we used the change of variable $y = x - 1$ and the final equality follows by recalling the relationship from equations (B.18) - (B.23) from the proof of the covariance matrix for the exact simulation scheme.

The remaining entries in Σ follow by symmetry.

B.3 Timer Options

Here we prove the fair price of a *timer option* in an economy with zero interest rate and dividend yield.¹

Assume therefore an asset with price process $(Z_t)_{t \geq 0}$ whose risk-neutral dynamics are given by

$$dZ_t = Z_t \sqrt{V_t^z} dW_{z,t} \quad (\text{B.61})$$

$$dV_t^z = \alpha_t dt + \beta_t dW_{v,t} \quad (\text{B.62})$$

where $W_{z,t}$ and $W_{v,t}$ are two correlated Brownian motions and α_t, β_t are well-behaved stochastic processes that are measurable with respect to the natural filtration generated by the market and that ensures $V_t^z > 0$ almost surely and that $\Theta_t := \int_0^t V_u^z du$ exists and converges to infinity for $t \rightarrow \infty$.

A timer option on Z with log-strike k and variance budget $V > 0$ is then an option that pays

$$(Z_{\tau_V} - e^k)^+ \quad (\text{B.63})$$

at the random time point $\tau_V := \inf\{t > 0 : \Theta_t = V\}$.

Let us now consider the problem of pricing this option.

First off, a quick application of Ito shows that

$$\log Z_t = \log Z_0 + \int_0^t \sqrt{V_u^z} dW_{z,u} - \frac{1}{2} \Theta_t. \quad (\text{B.64})$$

Recall now briefly the Dubins-Schwartz theorem:

Theorem B.3.1 (Dubins-Schwartz). *Let $(M_t)_{t \geq 0}$ be a continuous martingale with $M_0 = 0$ and $\langle M \rangle_\infty = +\infty$. Then there exists a Brownian motion $(B_t)_{t \geq 0}$ st. for every $t \geq 0$:*

$$M_t = B_{\langle M \rangle_t}. \quad (\text{B.65})$$

By the above there then exists a Brownian motion $(B_t)_{t \geq 0}$ such that

$$\log Z_t = \log Z_0 + B_{\Theta_t} - \frac{1}{2} \Theta_t \quad (\text{B.66})$$

which then implies

$$Z_{\tau_V} = Z_0 e^{B_{\Theta_{\tau_V}} - \frac{1}{2} \Theta_{\tau_V}} = Z_0 e^{B_V - \frac{1}{2} V} \quad (\text{B.67})$$

as τ_V is exactly the time point at which $\Theta_{\tau_V} = V$.

Let now $\text{BS}(\nu; z, k)$ denote the Black-Scholes formula when inputting total variance ν , spot z , log strike k and keeping interest rates and dividends at zero.²

Combining (B.63) and (B.67) the time zero price of the timer option must then be $\text{BS}(V; Z_0, k)$.

¹For a more general exposition of timer options see (Bernard & Cui 2010).

²By *total variance* we mean the product of the Black-Scholes squared volatility and time-to-expiry

B.4 Functional Derivatives

B.4.1 The Concept

Here we explain briefly the concept of a functional and a functional derivative in the context of real-valued functions with domain \mathbb{R} or subsets thereof. Our exposition draws inspiration from the explanation in appendix A in (Engel & Dreizler 2011). For a more mathematically rigorous explanation the reader will have to consult other works in the literature.

A functional F is a mapping from a space of functions to the real numbers. If $f : \mathbb{R} \rightarrow \mathbb{R}$ is a function then we use the notation $F[f]$ for the functional F evaluated in the function f .³

The concept of a derivative of a functional is not contained in the standard notions of derivatives for ordinary functions. Instead we need the concept of a so-called functional derivative.

In the following we let f denote some fixed function in the domain of F . We now wish to investigate what happens to F locally around f . In particular we consider the change in F when we modify f a little. Let therefore $\eta : \mathbb{R} \rightarrow \mathbb{R}$ be a so-called test function. Then we consider the change $\delta f := \epsilon\eta$ for some small $\epsilon \in \mathbb{R}$ which is then also a function on the same space as f and η .

We can now define the variation in F , denoted δF , as

$$\delta F := F[f + \delta f] - F[f] = F[f + \epsilon\eta] - F[f]. \quad (\text{B.68})$$

Since we can consider δF as an ordinary function of ϵ we can now use a standard Taylor expansion around $\epsilon = 0$ to get the approximation:⁴

$$F[f + \epsilon\eta] = F[f] + \frac{dF[f + \epsilon\eta]}{d\epsilon}|_{\epsilon=0}\epsilon + \frac{1}{2} \frac{d^2F[f + \epsilon\eta]}{d\epsilon^2}|_{\epsilon=0}\epsilon^2 + \dots \quad (\text{B.69})$$

We now define the functional derivative $\frac{\delta F[f]}{\delta f} : \mathbb{R} \rightarrow \mathbb{R}$ implicitly as the function that satisfies⁵

$$\int_{\mathbb{R}} \frac{\delta F[f]}{\delta f(x_1)} \eta(x_1) dx_1 := \frac{dF[f + \epsilon\eta]}{d\epsilon}|_{\epsilon=0}. \quad (\text{B.70})$$

There is no guarantee that any function f , functional F and test function η allows the representation above. However, in case it is possible we say that the functional F is differentiable.

In a similar fashion we define implicitly also the second order functional derivative $\frac{\delta^2 F[f]}{\delta f \delta f} : \mathbb{R}^2 \rightarrow \mathbb{R}$ via the relation

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\delta^2 F[f]}{\delta f(x_1) \delta f(x_2)} \eta(x_1) \eta(x_2) dx_1 dx_2 := \frac{d^2 F[f + \epsilon\eta]}{d\epsilon^2}|_{\epsilon=0}. \quad (\text{B.71})$$

The definition generalizes in the obvious way for higher order functional derivatives. In particular we define $\frac{\delta^n F[f]}{\delta f \dots \delta f} : \mathbb{R}^n \rightarrow \mathbb{R}$ implicitly through the relation

$$\int_{\mathbb{R}} \dots \int_{\mathbb{R}} \frac{\delta^n F[f]}{\delta f(x_1) \dots \delta f(x_n)} \eta(x_1) \dots \eta(x_n) dx_1 \dots dx_n := \frac{d^n F[f + \epsilon\eta]}{d\epsilon^n}|_{\epsilon=0}. \quad (\text{B.72})$$

³The concepts below can easily be generalized if the domain of f is some subset of the real line.

⁴The approximation is naturally only valid if $\epsilon \mapsto F[f + \epsilon\eta]$ is sufficiently differentiable at $\epsilon = 0$.

⁵A natural notation for evaluating $\frac{\delta F[f]}{\delta f}$ at x would be $\frac{\delta F[f]}{\delta f}(x)$. However, as it is common in the literature to write $\frac{\delta F[f]}{\delta f(x)}$ instead we stick to this convention.

To gain some intuition on what the concept of a functional derivative means consider a function $g : \mathbb{R}^N \rightarrow \mathbb{R}$ of finitely many variables. A Taylor expansion of first order around a point $(x_1^0, \dots, x_N^0) \in \mathbb{R}^N$ then gives

$$g(x_1, \dots, x_N) \approx g(x_1^0, \dots, x_N^0) + \sum_{i=1}^N \frac{\partial g}{\partial x_i}(x_1^0, \dots, x_N^0) \cdot (x_i - x_i^0). \quad (\text{B.73})$$

Comparing the above to B.69 and B.70 we see that the functional derivative $\frac{\delta F[f]}{\delta f}$ generalizes the notion of a partial derivative to functions of infinitely many variables, i.e. functionals.

The intuition works similarly for higher-order functional derivatives.

B.4.2 Numerical Computation of First Order Functional Derivatives

In this subsection we consider how to compute first order functional derivatives of the form $\frac{\delta F[f]}{\delta f}$ numerically. This is useful when we don't know the functional F analytically.

For simplicity we assume the domain of f is the positive real line only. Consider first a discretization of the domain of f given by $t_i = \frac{i}{n}$ for $i = 0, 1, \dots$ where we have also chosen some large $n \in \mathbb{N}$. Now to estimate $\frac{\delta F[f]}{\delta f(t_j)}$ for some $j \in \mathbb{N}$ we simply consider the test function $\eta_j(x) = 1$ for $x \in [t_{j-1}, t_j]$ and $\eta_j(x) = 0$ otherwise. We now have

$$\frac{dF[f + \epsilon \eta_j]}{d\epsilon} \Big|_{\epsilon=0} \stackrel{\text{def.}}{=} \int_{\mathbb{R}_+} \frac{\delta F[f]}{\delta f(x)} \eta_j(x) dx \quad (\text{B.74})$$

$$= \sum_{i=1}^{\infty} \int_{t_{i-1}}^{t_i} \frac{\delta F[f]}{\delta f(x)} \eta_j(x) dx \quad (\text{B.75})$$

$$= \int_{t_{j-1}}^{t_j} \frac{\delta F[f]}{\delta f(x)} dx \quad (\text{B.76})$$

$$\approx (t_j - t_{j-1}) \frac{\delta F[f]}{\delta f(t_j)} \quad (\text{B.77})$$

$$(\text{B.78})$$

and thus we get

$$\frac{\delta F[f]}{\delta f(t_j)} \approx (t_j - t_{j-1})^{-1} \frac{dF[f + \epsilon \eta_j]}{d\epsilon} \Big|_{\epsilon=0} \quad (\text{B.79})$$

where the ordinary derivative on the right hand side can be computed using finite difference methods.

B.5 Forward Variance Swaps

Per equation (5.24) we already know how to price variance swaps. In some cases they are even traded actively. However, here we wish to illustrate how one can also trade the forward variances in an appropriate sense. In the following we assume the interest rate $r(t) = r$ is constant. The arguments can easily be extended to the more general case of a time dependent but deterministic interest rate.

Recall first the relation

$$\xi_t(u) = \frac{d}{dT} ((T-t)\hat{\sigma}_{VS,T}^2(t))|_{T=u} \quad (\text{B.80})$$

where we the maturity T implied variance swap volatility observed at $t < T$ is given as

$$\hat{\sigma}_{VS,T}(t) = \sqrt{\frac{1}{T-t} E^Q \left(\int_t^T \sigma_s^2 ds \middle| \mathcal{F}_t \right)}. \quad (\text{B.81})$$

The time T payoff of such a variance swap, bought at time t , is

$$\int_t^T \sigma_s^2 ds - (T-t)\hat{\sigma}_{VS,T}^2(t). \quad (\text{B.82})$$

Let the current time point be $t > 0$. The goal now is to construct a zero-cost self-financing trading strategy that produces a per period PnL of $d\xi_t(u)$ for some forward variance maturity $u > t$.

Consider two time points T_1, T_2 close to u such that $T_2 > u > T_1 > t$.

Say first that we buy 1 maturity T_2 variance swap at an implied volatility of $\hat{\sigma}_{VS,T_2}(t)$ and sell $e^{-r(T_2-T_1)}$ maturity T_1 variance swaps at an implied volatility of $\hat{\sigma}_{VS,T_1}(t)$. As we are trading variance swaps the cost of forming this portfolio is zero. A time T_1 we receive the pay-off of our T_1 variance swap position. We immediately place the gain or loss in the risk-free asset. At time T_2 we receive the payoff of the T_2 variance swap and unwind our entire portfolio.

The value of this self-financing portfolio strategy at time T_2 is

$$\int_t^{T_2} \sigma_s^2 ds - (T_2-t)\hat{\sigma}_{VS,T_2}^2(t) - e^{r(T_2-T_1)} \cdot e^{-r(T_2-T_1)} \cdot \left(\int_t^{T_1} \sigma_s^2 ds - (T_1-t)\hat{\sigma}_{VS,T_1}^2(t) \right) \quad (\text{B.83})$$

$$= \int_{T_1}^{T_2} \sigma_s^2 ds - ((T_2-t)\hat{\sigma}_{VS,T_2}^2(t) - (T_1-t)\hat{\sigma}_{VS,T_1}^2(t)) \quad (\text{B.84})$$

$$= \int_{T_1}^{T_2} \sigma_s^2 ds - (T_2-T_1)\hat{\sigma}_{VS,T_1,T_2}^2(t) \quad (\text{B.85})$$

where we have defined the discrete forward variance $\hat{\sigma}_{VS,T_1,T_2}(t)$ as

$$\hat{\sigma}_{VS,T_1,T_2}^2(t) = \frac{(T_2-t)\hat{\sigma}_{VS,T_2}^2(t) - (T_1-t)\hat{\sigma}_{VS,T_1}^2(t)}{T_2-T_1}. \quad (\text{B.86})$$

For T_1, T_2 close to u the above is a discrete approximation to the derivative $\frac{d}{dT} (f(T))|_{T=u}$ where $f(T) := (T-t)\hat{\sigma}_{VS,T}^2(t)$ and thus we conclude $\hat{\sigma}_{VS,T_1,T_2}^2(t) \approx \xi_t(u)$.

We now consider a different variation of the above strategy. To this end we take a time point t' such that $t < t' < T_1$. We think of t' being close to t so $t' - t \approx dt$ is a small time step.

For this strategy we at time t' sell 1 maturity T_2 swap and buy $e^{-r(T_2-T_1)}$ maturity T_1 swaps. At T_1 we receive the pay-off of the one variance swap. We place any proceeds in the bank account at the rate r . The time T_2 PnL of this other portfolio strategy is then

$$-\left(\int_{t'}^{T_2} \sigma_s^2 ds - (T_2 - t')\hat{\sigma}_{VS,T_2}^2(t')\right) + e^{r(T_2-T_1)} \cdot e^{-r(T_2-T_1)} \cdot \left(\int_{t'}^{T_1} \sigma_s^2 ds - (T_1 - t')\hat{\sigma}_{VS,T_1}^2(t')\right) \quad (\text{B.87})$$

$$= -\int_{T_1}^{T_2} \sigma_s^2 ds + (T_2 - t')\hat{\sigma}_{VS,T_2}^2(t') - (T_1 - t')\hat{\sigma}_{VS,T_1}^2(t') \quad (\text{B.88})$$

$$= -\int_{T_1}^{T_2} \sigma_s^2 ds + (T_2 - T_1)\hat{\sigma}_{VS,T_1,T_2}^2(t'). \quad (\text{B.89})$$

The combined portfolio resulting from trading both these strategies then results in a PnL of exactly

$$(T_2 - T_1) (\hat{\sigma}_{VS,T_1,T_2}^2(t') - \hat{\sigma}_{VS,T_1,T_2}^2(t)) \quad (\text{B.90})$$

at time T_2 .

As this is a known quantity at t' the value of this portfolio at time t' is exactly

$$e^{-r(T_2-t')} \cdot (T_2 - T_1) \cdot (\hat{\sigma}_{VS,T_1,T_2}^2(t') - \hat{\sigma}_{VS,T_1,T_2}^2(t)). \quad (\text{B.91})$$

Thus by scaling the combined trading strategy by $e^{r(T_2-t')}(T_2 - T_1)^{-1}$ units we get exactly a PnL from time t to time t' of

$$\hat{\sigma}_{VS,T_1,T_2}^2(t') - \hat{\sigma}_{VS,T_1,T_2}^2(t) \approx \xi_{t'}(u) - \xi_t(u). \quad (\text{B.92})$$

We conclude that it is possible, with an appropriate trading strategy in variance swaps, to get a per period PnL of $d\xi_t(u)$ and that at zero cost. We refer to the contract that this construction implies as a *forward variance swap*.

Since the forward variance swap costs nothing but gives a per period PnL of $d\xi_t(u)$ it must hold that $\xi_t(u)$ is a martingale under the pricing measure Q to avoid arbitrage. Assuming also that $\xi_t(u)$ is a continuous process it must have dynamics of the form $d\xi_t(u) = \dots dW_t^Q$ where W_t^Q is some Brownian motion under the pricing measure Q ⁶.

⁶Here we use the Martingale Representation Theorem.

B.6 The Bergomi-Guyon Expansion Formula

In (Bergomi & Guyon 2011) Lorenzo Bergomi and Julien Guyon derives an expansion formula for implied volatilities in a general class of multi-factor diffusive stochastic volatility models. Here we briefly recite the results of their work.

Assume a modelling framework with dynamics under the pricing measure Q as

$$d \log(S_t) = -\frac{1}{2} \xi_t(t) dt + \sqrt{\xi_t(t)} dZ_t^1 \quad (\text{B.93})$$

$$d\xi_t(u) = \lambda(t, u, \xi_t) \cdot dZ_t \quad (\text{B.94})$$

where $\lambda = (\lambda_1, \dots, \lambda_d)$ takes values in \mathbb{R}^d and $Z = (Z^1, \dots, Z^d)$ is a d -dimensional Brownian motion where only the first component Z^1 drives the spot process. For simplicity we have assumed no dividends or interest rates.

Then at second order in ϵ , the implied volatility for a expiry T , strike K vanilla option is⁷

$$\hat{\sigma}(T, K)^\epsilon = \hat{\sigma}_T^{ATM} + \mathcal{S}_T \ln \left(\frac{K}{S_0} \right) + \mathcal{C}_T \ln^2 \left(\frac{K}{S_0} \right) + O(\epsilon^2) \quad (\text{B.95})$$

where

$$\hat{\sigma}_T^{ATM} = \hat{\sigma}_T^{VS} \left[1 + \frac{\epsilon}{4v} C^{x\xi} + \frac{\epsilon^2}{32v^3} (12(C^{x\xi})^2 - v(v+4)C^{\xi\xi} + 4v(v-4)C^\mu) \right] \quad (\text{B.96})$$

$$\mathcal{S}_T = \hat{\sigma}_T^{VS} \left[\frac{\epsilon}{2v^2} C^{x\xi} + \frac{\epsilon^2}{8v^3} (4C^\mu v - 3(C^{x\xi})^2) \right] \quad (\text{B.97})$$

$$\mathcal{C}_T = \hat{\sigma}_T^{VS} \frac{\epsilon^2}{8v^4} (4C^\mu v + C^{\xi\xi} v - 6(C^{x\xi})^2) \quad (\text{B.98})$$

with $v = \int_0^T \xi_0(t) dt$, $\hat{\sigma}_T^{VS} = \sqrt{\frac{v}{T}}$ being the variance swap implied volatility for maturity T and the constants $C^{x\xi}$, $C^{\xi\xi}$ and C^μ to be defined in a moment. For numerical applications one should put $\epsilon = 1$. Also in a setting with non-zero but deterministic interest rate and continuous and deterministic dividend yield the formula still works by replacing the spot price S_0 by its forward price.⁸

A quick differentiation will reveal that at order two in ϵ , the at-the-money skew is given by \mathcal{S}_T .

Recalling the definitions of μ and ν from equations (6.19) and (6.20) we define the integrated spot/variance covariance function as

$$C_t^{x\xi}(\xi) = \int_t^T \int_s^T \mu(s, u, \xi) du ds, \quad (\text{B.99})$$

the integrated variance/variance covariance function as

$$C_t^{\xi\xi}(\xi) = \int_t^T \int_s^T \int_s^T \nu(s, u, u', \xi) du' du ds \quad (\text{B.100})$$

and finally we define

$$C_t^\mu(\xi) = \int_t^T \int_s^T \mu(s, u, \xi) \frac{\delta C_t^{x\xi}}{\delta \xi_t(u)} du ds. \quad (\text{B.101})$$

⁷For details about in which way the expansion in ϵ is to be understood we refer the reader to the original work in (Bergomi & Guyon 2011).

⁸This is easily shown by writing up the definition of the Black-Scholes implied volatility.

In the expansion formula from above we then use the definitions $C^{x\xi} := C_0^{x\xi}(\xi_0)$, $C^{\xi\xi} := C_0^{\xi\xi}(\xi_0)$ and $C^\mu := C_0^\mu(\xi_0)$.

Explicit formulas for these constants are derived in (Bayer et al. 2016) for the case of the rough Bergomi model. They simplify somewhat for a flat forward variance curve.

In particular we get

$$C^{x\xi} = \int_0^T \int_t^T \frac{E(d\log(S_t)d\xi_t(u))}{dt} du dt \quad (\text{B.102})$$

$$= \rho\tilde{\eta}\xi_0(0)^{3/2} \int_0^T \int_t^T (u-t)^{-\gamma} du dt \quad (\text{B.103})$$

$$= \rho\tilde{\eta}\xi_0(0)^{3/2} \int_0^T \frac{1}{1-\gamma} (T-t)^{1-\gamma} dt \quad (\text{B.104})$$

$$= \frac{\rho\tilde{\eta}\xi_0(0)^{3/2}}{(1-\gamma)(2-\gamma)} T^{2-\gamma}. \quad (\text{B.105})$$

Thus at order one in ϵ we have the following skew approximation under the rough Bergomi model:

$$\psi(T) = \xi_0(0)^{1/2} \frac{1}{2T^2\xi_0(0)^2} C^{x\xi} \quad (\text{B.106})$$

$$= \xi_0(0)^{1/2} \frac{1}{2T^2\xi_0(0)^2} \frac{\rho\tilde{\eta}\xi_0(0)^{3/2}}{(1-\gamma)(2-\gamma)} T^{2-\gamma} \quad (\text{B.107})$$

$$= \frac{\rho\tilde{\eta}}{2} [(1-\gamma)(2-\gamma)]^{-1} T^{-\gamma}. \quad (\text{B.108})$$

For a flat forward variance curve (Bayer et al. 2016) furthermore shows that the rough Bergomi model to an approximation leads to the terms

$$C^{\xi\xi} = \eta^2\xi_0(0)^2 D_H^2 \frac{T^{2+2H}}{2+2H} \quad (\text{B.109})$$

and

$$C^\mu = \frac{1}{2} \rho^2 \eta^2 \xi_0(0)^2 D_H^2 \left(1 + \frac{\Gamma(H+3/2)^2}{\Gamma(2H+3)} \right) \frac{T^{2+2H}}{2+2H}. \quad (\text{B.110})$$

Using these in (B.97) and letting $\epsilon = 1$ we get a second order Bergomi-Guyon approximation of skew for a flat forward variance curve under the rough Bergomi model.

B.7 The Skew-Stickiness-Ratio

Here we present and prove some results related to the skew-stickiness-ratio (abbreviated SSR). Chapter 9 in (Bergomi 2016) has been a large source of inspiration for this note.

Recall that the SSR for maturity $T > 0$ is defined as

$$R_T := \frac{1}{\psi(T)} \frac{E(d \log S d\hat{\sigma}_{F_T})}{E((d \log S)^2)} \quad (\text{B.111})$$

where $\hat{\sigma}_{F_T}$ is the at-the-money, as in $k = 0$, implied volatility for a maturity T vanilla option.⁹

A brief remark: In the following we will be making a decent amount of approximations. For that purpose we will to some extend abuse notation and for instance write ' $=$ ' when really it should be ' \approx '. We will however always make it explicit what approximations are used in each step.

B.7.1 Approximate expressions

Here we will derive some more explicit, though approximate, expressions for the SSR as well as at-the-money skew based on the Bergomi-Guyon expansion from section B.6. We will assume that the given model we are working with can be written on forward variance curve form.

From the Bergomi-Guyon expansion formula the difference between the at-the-money and variance swap (VS) implied volatility is of order one. To an approximation we can therefore replace $\hat{\sigma}_{F_T}$ by the VS volatility $\hat{\sigma}_T^{VS}$.

Let us now compute $E[d \ln S d\hat{\sigma}_T^{VS}]$.

Note first that $\hat{\sigma}_T^{VS}(t) = \sqrt{\frac{1}{T-t} \int_t^T \xi_t(u) du}$ such that

$$d\hat{\sigma}_T^{VS}(t) = ... dt + \int_t^T \frac{1}{2} (\hat{\sigma}_T^{VS}(t))^{-1} \frac{1}{T-u} d\xi_t(u) du = ... dt + \frac{1}{2(T-t)\hat{\sigma}_T^{VS}(t)} \int_t^T d\xi_t(u) du \quad (\text{B.112})$$

and thus

$$E[d \ln S_t d\hat{\sigma}_T^{VS}(t)] = \frac{1}{2(T-t)\hat{\sigma}_T^{VS}(t)} \int_t^T E[d \ln S_t d\xi_t(u)] du \quad (\text{B.113})$$

$$= \frac{1}{2(T-t)\hat{\sigma}_T^{VS}(t)} \int_t^T \mu(t, u, \xi_t) du dt. \quad (\text{B.114})$$

Assume now for simplicity that $t = 0$.

Using the Bergomi-Guyon expansion formula for skew at order one with $\epsilon = 1$ we therefore get

$$\psi(T) = \frac{1}{2\sqrt{T}} \frac{1}{\left(\int_0^T \xi_0(\tau) d\tau\right)^{3/2}} \int_0^T \int_\tau^T \mu(\tau, u, \xi_0) du d\tau. \quad (\text{B.115})$$

⁹We use the extra subscript F_T to stress that since we consider the at-the-money implied volatility the strike is not fixed across time. Specifically $\hat{\sigma}_{F_T}$ is the implied volatility of a maturity T , strike F_T vanilla option where F_T is the forward price.

Noting further that $E((d \ln S)^2) = \xi_0(0)dt$ and combining all the above we get to an approximation

$$R_T = \frac{2\sqrt{T} \left(\int_0^T \xi_0(u) du \right)^{3/2}}{\int_0^T \int_\tau^T \mu(\tau, u, \xi_0) dud\tau} \cdot \frac{1}{2T \frac{1}{\sqrt{T}} \sqrt{\int_0^T \xi_0(u) du}} \int_0^T \mu(t, u, \xi_0) du \cdot \frac{1}{\xi_0(0)} \quad (\text{B.116})$$

$$= \frac{\int_0^T \xi_0(\tau) d\tau}{T \xi_0(0)} \frac{T \int_0^T \mu(0, u, \xi_0) du}{\int_0^T \int_\tau^T \mu(\tau, u, \xi_0) dud\tau}. \quad (\text{B.117})$$

We see that the SSR to an approximation only depends on the spot/variance covariance function μ .

We state the short maturity limit as a result:

Theorem B.7.1. *Under the assumptions and approximations introduced so far and assuming further that $u \mapsto \mu(t, u, \xi)$ as well as $u \mapsto \xi_t(u)$ are sufficiently differentiable in $u = t$ we should expect*

$$R_0 := \lim_{T \rightarrow 0} R_T = 2. \quad (\text{B.118})$$

Proof. A Taylor expansion around $T = 0$ yields

$$\int_0^T \mu(0, u, \xi_0) du = \mu(0, 0, \xi_0)T + \mathcal{O}(T^2) \quad (\text{B.119})$$

for the nominator and

$$\int_0^T \int_\tau^T \mu(\tau, u, \xi_0) dud\tau = \frac{1}{2} \mu(0, 0, \xi_0) T^2 + \mathcal{O}(T^3) \quad (\text{B.120})$$

for the denominator.

Similarly another Taylor expansion around $T = 0$ yields

$$\int_0^T \xi_0(\tau) d\tau = \xi_0(0)T + \mathcal{O}(T^2). \quad (\text{B.121})$$

Combining the above we get

$$R_0 := \lim_{T \rightarrow 0} R_T \quad (\text{B.122})$$

$$= \lim_{T \rightarrow 0} \frac{\int_0^T \xi_0(\tau) d\tau}{T \xi_0(0)} \frac{T \int_0^T \mu(0, u, \xi_0) du}{\int_0^T \int_\tau^T \mu(\tau, u, \xi_0) dud\tau} \quad (\text{B.123})$$

$$= \lim_{T \rightarrow 0} \frac{\xi_0(0)T + \mathcal{O}(T^2)}{\xi_0(0)T} \cdot \frac{\mu(0, 0, \xi_0)T^2 + \mathcal{O}(T^3)}{\frac{1}{2}\mu(0, 0, \xi_0)T^2 + \mathcal{O}(T^3)} \quad (\text{B.124})$$

$$= \lim_{T \rightarrow 0} \frac{\xi_0(0)T}{\xi_0(0)T} \cdot \frac{\mu(0, 0, \xi_0)T^2}{\frac{1}{2}\mu(0, 0, \xi_0)T^2} \quad (\text{B.125})$$

$$= 1 \cdot \frac{1}{\frac{1}{2}} \quad (\text{B.126})$$

$$= 2. \quad (\text{B.127})$$

for $T \rightarrow 0$. \square

The short maturity limit for skew is instead:

Theorem B.7.2. *Under the assumptions and approximations introduced so far and assuming further that $u \mapsto \mu(t, u, \xi)$ as well as $u \mapsto \xi_t(u)$ are sufficiently differentiable in $u = t$ we should expect*

$$\psi(0) := \lim_{T \rightarrow 0} \psi(T) = \frac{\mu(0, 0, \xi_0)}{4\xi_0(0)^{3/2}} \quad (\text{B.128})$$

i.e. short term skew converges to a constant.

Proof. Consider the approximate formula (B.115). Using two Taylor expansions around $T = 0$ we get

$$\psi(0) := \lim_{T \rightarrow 0} \psi(T) \quad (\text{B.129})$$

$$= \lim_{T \rightarrow 0} \frac{1}{2\sqrt{T}} \frac{1}{\left(\int_0^T \xi_0(\tau) d\tau\right)^{3/2}} \int_0^T \int_\tau^T \mu(\tau, u, \xi_0) du d\tau \quad (\text{B.130})$$

$$= \lim_{T \rightarrow 0} \frac{1}{2\sqrt{T}} \frac{1}{(\xi_0(0)T + \mathcal{O}(T^2))^{3/2}} \left(\frac{1}{2} \mu(0, 0, \xi_0) T^2 + \mathcal{O}(T^3) \right) \quad (\text{B.131})$$

$$= \frac{\mu(0, 0, \xi_0)}{4\xi_0(0)^{3/2}} \quad (\text{B.132})$$

which is constant. \square

It is important to note that the arguments used in the above two results does not work if we attempt to use it on the rough Bergomi model as the spot/variance covariance function $\mu(t, u, \xi_t) = \sqrt{\xi_t(t)} \xi_t(u) \tilde{\eta}(u-t)^{-\gamma}$ then is not even well-defined for $u = t$.

Assume now that the forward variance curve ξ_0 is flat. In this context we will abuse notation and simply write ξ_0 also for this fixed value. Assume furthermore that $\mu(t, u, \xi_0)$ only depends on $u - t$. To simplify we will now change notation to $\mu(u - t)$. Under these assumptions the expressions will simplify.

Defining $g(t) := \int_0^t \mu(\tau) d\tau$ we first off have:

$$\int_0^T \int_\tau^T \mu(u - \tau) du d\tau = \int_0^T \int_0^{T-t} \mu(u) du dt \quad (\text{B.133})$$

$$= \int_0^T g(T-t) dt \quad (\text{B.134})$$

$$= \int_0^T g(t) dt \quad (\text{B.135})$$

$$= [g(t)(t-T)]_0^T - \int_0^T g'(t)(t-T) dt \quad (\text{B.136})$$

$$= \int_0^T \mu(t)(T-t) dt. \quad (\text{B.137})$$

Combining (B.115) and (B.117) with (B.137) and using our extra assumptions we now get the simpler expressions

$$\psi(T) = \frac{\int_0^T (T-t) \mu(t) dt}{2(\xi_0)^{3/2} T^2} \quad (\text{B.138})$$

and

$$R_T = \frac{\int_0^T \mu(t) dt}{\int_0^T \mu(t)(1 - \frac{t}{T}) dt}. \quad (\text{B.139})$$

We now want to consider the admissible range for R_T . For that purpose we need to rephrase the above expressions in terms of g . Thus note that

$$\psi(T) = \frac{1}{2\xi_0^{3/2} T^2} \int_0^T g(\tau) d\tau \quad (\text{B.140})$$

and

$$R_T = \frac{g(T)}{\frac{1}{T} \int_0^T g(\tau) d\tau}. \quad (\text{B.141})$$

Theorem B.7.3. *Under the assumptions and approximations introduced so far and assuming further that $\mu(t)$ is monotonically decaying towards 0 for $t \rightarrow \infty$ we should expect $R_T \in [1, 2]$.*

Proof. Since $\mu(t)$ is monotonically decaying towards 0 for $t \rightarrow \infty$ we conclude that $g(T) = \int_0^T \mu(t) dt$ is either monotonically increasing or decreasing with $g(0) = 0$. Therefore $|g(T)| \geq |\frac{1}{T} \int_0^T g(\tau) d\tau|$. It also holds that $g(T)$ and its average have the same sign. This allows us to conclude $R_T \geq 1$.

Say now that $\mu(t)$ is decreasing from above zero. Since $g'(t) = \mu(t)$, $g''(t) = \mu'(t)$ and in this case $\mu'(t)$ is negative, we then know that g is increasing and concave. In case $\mu(t)$ is increasing from below zero we instead get that g is decreasing and convex.

We can now argue $\frac{g(\tau)}{g(T)} \geq \frac{\tau}{T}$ for $\tau < T$. To see this consider the case of g being positive and concave. Then $\tau \mapsto g(\tau)$ increases faster than linearly towards $g(T)$ in the beginning. Thus the fraction achieved at τ , that is $\frac{g(\tau)}{g(T)}$, is higher than the fraction of the time distance travelled, that is $\frac{\tau}{T}$. A similar argument applies in the case of g being negative and convex.

In total we can conclude:

$$R_T = \frac{1}{\frac{1}{T} \int_0^T \frac{g(\tau)}{g(T)} d\tau} \leq \frac{1}{\frac{1}{T} \int_0^T \frac{\tau}{T} d\tau} = 2. \quad (\text{B.142})$$

□

Thus for a time-homogeneous monotonically decaying spot/variance covariance function, and to the approximation of a flat forward variance curve, we should expect $R_T \in [1, 2]$. However, under a general non-flat forward variance curve we should expect SSR values outside the range $[1, 2]$ to be possible.

B.7.2 Type I and Type II models

We now wish to say a little bit more about how skew and stickiness behaves for $T \rightarrow \infty$. To this end we maintain the assumption of a time-homogeneous spot/variance covariance function and further assume $\mu(t) \sim t^{-\gamma}$ for $t \rightarrow \infty$.¹⁰

Recall in the following our definition $g(\tau) := \int_0^\tau \mu(t) dt$ as well as the representation in (B.141).

¹⁰Here the symbol '∞' simply means that the ratio of the left- and right-hand-side converge to a constant, i.e. are roughly proportional for large t .

For large τ we then have that $g(\tau)$ is approximately equal to $C + \alpha\tau^{1-\gamma}$ where C and α are constants. Thus if $\gamma > 1$ then $g(\tau)$ converges to C . If instead $\gamma < 1$ then $g(\tau)$ behaves as $\alpha\tau^{1-\gamma}$ for large τ .

Note also that $\int_0^T g(\tau)d\tau$ for large T behaves as

$$B + CT + \frac{\alpha}{2-\gamma} T^{2-\gamma} \quad (\text{B.143})$$

where B is another constant.

Therefore if $\gamma > 1$ we see that $\frac{1}{T} \int_0^T g(\tau)d\tau$ for large T essentially is equal to C . Consulting (B.141) we in this case therefore get $R_\infty := \lim_{T \rightarrow \infty} R_T = 1$. In case $\gamma < 1$ we instead get that the average behaves as $\frac{\alpha}{2-\gamma} T^{1-\gamma}$ for large enough T . In this case we conclude $R_\infty = 2 - \gamma$.

Looking at the skew expression (B.140) we also have $\psi(T) \sim T^{-1}$ if $\gamma > 1$ since (B.143) then behaves as CT for large T and if instead $\gamma < 1$ then it behaves as $\frac{\alpha}{2-\gamma} T^{2-\gamma}$ such that we can conclude $\psi(T) \sim T^{-\gamma}$.

In total we have identified two types of models depending on γ :

Type I: If $\gamma > 1$ then $\psi(T) \sim T^{-1}$ for large T and $R_\infty = 1$.

Type II: If $\gamma < 1$ then $\psi(T) \sim T^{-\gamma}$ for large T and $R_\infty = 2 - \gamma$.

In essence, under the assumptions introduced so far, the rate of decay of skew and the value of the SSR are really two sides of the same coin and are both related to the rate of decay of the spot-variance covariance function $\mu(t)$. We do however remark that if e.g. $\mu(t)$ has an exponential form then it exhibits exactly the properties of a type I model. We prove this in the next subsection.

B.7.3 Exponential Kernels

In the definition of type I and II models the spot/variance covariance function must decay as a power-law. Here we instead assume an exponential kernel and show that it behaves as a type I model. We present the results in the following theorem:

Theorem B.7.4. *Assume*

$$\mu(t) \sim \sum_{i=1}^n \eta_i e^{-\kappa_i t} \quad (\text{B.144})$$

for $t \rightarrow \infty$ and $\eta_i, \kappa_i > 0$, $i = 1, \dots, n$, $n \in \mathbb{N}$ being constants.

Then we should expect $R_\infty = 1$, $\psi(T) \sim \frac{1}{T}$ for $T \rightarrow \infty$.

Proof. For simplicity of the derivation we assume (B.144) holds with equality for large enough t .

Note first that

$$g(\tau) = \int_0^\tau \mu(t) dt \quad (\text{B.145})$$

$$= \sum_{i=1}^n \eta_i \int_0^\tau e^{-\kappa_i t} dt \quad (\text{B.146})$$

$$= \sum_{i=1}^n \eta_i \cdot \frac{1}{-\kappa_i} (e^{-\kappa_i \tau} - 1) \quad (\text{B.147})$$

$$= \sum_{i=1}^n \frac{\eta_i}{\kappa_i} (1 - e^{-\kappa_i \tau}) \quad (\text{B.148})$$

such that also

$$\int_0^T g(\tau) d\tau = \sum_{i=1}^n \frac{\eta_i}{\kappa_i} \int_0^T (1 - e^{-\kappa_i \tau}) d\tau \quad (\text{B.149})$$

$$= T \sum_{i=1}^n \frac{\eta_i}{\kappa_i} - \sum_{i=1}^n \frac{\eta_i}{\kappa_i} \int_0^T e^{-\kappa_i \tau} d\tau \quad (\text{B.150})$$

$$= T \sum_{i=1}^n \frac{\eta_i}{\kappa_i} - \sum_{i=1}^n \frac{\eta_i}{\kappa_i} \left(\frac{1}{-\kappa_i} (e^{-\kappa_i T} - 1) \right) \quad (\text{B.151})$$

$$= T \sum_{i=1}^n \frac{\eta_i}{\kappa_i} + \sum_{i=1}^n \frac{\eta_i}{\kappa_i^2} (e^{-\kappa_i T} - 1). \quad (\text{B.152})$$

Therefore

$$R_T = \frac{g(T)}{\frac{1}{T} \int_0^T g(\tau) d\tau} \quad (\text{B.153})$$

$$= \frac{\sum_{i=1}^n \frac{\eta_i}{\kappa_i} (1 - e^{-\kappa_i T})}{\sum_{i=1}^n \frac{\eta_i}{\kappa_i} + \frac{1}{T} \sum_{i=1}^n \frac{\eta_i}{\kappa_i^2} (e^{-\kappa_i T} - 1)} \quad (\text{B.154})$$

Letting $T \rightarrow \infty$ we see that

$$R_T \rightarrow \frac{\sum_{i=1}^n \frac{\eta_i}{\kappa_i}}{\sum_{i=1}^n \frac{\eta_i}{\kappa_i}} = 1. \quad (\text{B.155})$$

For the long maturity limit of skew we can simply consider the more specific expression in (B.140) and the previously derived expression for $\int_0^T g(\tau) d\tau$ to get:

$$\psi(T) \sim T^{-2} \int_0^T g(\tau) d\tau = T^{-2} \left(T \sum_{i=1}^n \frac{\eta_i}{\kappa_i} + \sum_{i=1}^n \frac{\eta_i}{\kappa_i^2} (e^{-\kappa_i T} - 1) \right) \quad (\text{B.156})$$

$$= \frac{1}{T} \sum_{i=1}^n \frac{\eta_i}{\kappa_i} + \frac{1}{T^2} \sum_{i=1}^n \frac{\eta_i}{\kappa_i^2} (e^{-\kappa_i T} - 1) \quad (\text{B.157})$$

In the limit this clearly behaves as T^{-1} and so we get $\psi(T) \sim \frac{1}{T}$ for $T \rightarrow \infty$.

□

Let us now consider some specific models that are well-known in the literature:

A general n -factor Bergomi forward variance curve model is defined under the pricing measure as

$$d\xi_t(u) = \xi_t(u) \left(\sum_{i=1}^n \eta_i e^{-\kappa_i(t-s)} dW_s^{(i)} \right). \quad (\text{B.158})$$

where $\eta_i, \kappa_i > 0$, $i = 1, \dots, n$ are parameters and $(W_s^{(i)})_{i=1}^n$ are correlated Brownian motions under the pricing measure. Assuming the dynamics of the stock price has the usual form and to the approximation of a flat forward variance curve we see that this model fits into the assumptions of theorem B.7.4. The exponential kernel thus causes the model to behave as a type I model.

Finally let us consider the Heston model. Assuming for simplicity zero interest rate and dividends, the Heston model is typically defined as

$$dS_t = \sqrt{v_t} S_t dW_{2,t} \quad (\text{B.159})$$

$$dv_t = -\kappa(v_t - \bar{v}) dt + \sigma \sqrt{v_t} dW_{1,t} \quad (\text{B.160})$$

where W_1, W_2 are two correlated Brownian motions and κ, σ, \bar{v} are parameters of the model.

Recalling that $\xi_t(u) = E^Q \left(v_u \middle| \mathcal{F}_t \right)$ for $u > t$ we find that

$$\frac{d\xi_t(u)}{du} = -\kappa(\xi_t(u) - \bar{v}) \quad (\text{B.161})$$

which is an ordinary differential equation with the solution

$$\xi_t(u) = \bar{v} + e^{-\kappa(u-t)} (\xi_t(t) - \bar{v}). \quad (\text{B.162})$$

An application of Ito's lemma on $\xi_t(u)$, recalling the dynamics of $\xi_t(t) = v_t$, then yields:

$$d\xi_t(u) = \kappa e^{-\kappa(u-t)} (\xi_t(t) - \bar{v}) dt + e^{-\kappa(u-t)} d\xi_t(t) \quad (\text{B.163})$$

$$= \kappa e^{-\kappa(u-t)} (\xi_t(t) - \bar{v}) dt + e^{-\kappa(u-t)} \left(-\kappa(\xi_t(t) - \bar{v}) dt + \sigma \sqrt{\xi_t(t)} dW_{1,t} \right) \quad (\text{B.164})$$

$$= \sigma e^{-\kappa(u-t)} \sqrt{\xi_t(t)} dW_{1,t}. \quad (\text{B.165})$$

Noticing the exponential form in the above dynamics we should thus similarly expect Heston to behave as a type I model under the extra approximation of a flat forward variance curve.

B.7.4 Computing the SSR numerically

In this section we detail how to compute the skew-stickiness-ratio numerically for a model without jumps that is Markov in the stock price S_t and the entire forward variance curve ξ_t .

Note that under these assumptions we must have that the expiry $T > 0$ at-the-money implied volatility $\hat{\sigma}_{FT}(t) =: \hat{\sigma}_{FT}(t, S_t, \xi_t)$ exactly is a function of these two components and time t .

An application of the infinite-dimensional version of Ito's lemma therefore yields

$$d\hat{\sigma}_{FT}(t) = \dots dt + \frac{d\hat{\sigma}_{FT}}{d \ln S} d \ln(S_t) + \int_t^T \frac{\delta \hat{\sigma}_{FT}}{\delta \xi_t(u)} d\xi_t(u) du \quad (\text{B.166})$$

$$= \dots dt + \frac{d\hat{\sigma}_{FT}}{d \ln S} \sqrt{\xi_t(t)} dW_{2,t} + \int_t^T \frac{\delta \hat{\sigma}_{FT}}{\delta \xi_t(u)} \xi_t(u) \tilde{\eta}(u-t)^{-\gamma} du dW_{1,t} \quad (\text{B.167})$$

$$= \dots dt + \frac{d\hat{\sigma}_{FT}}{d \ln S} \sqrt{\xi_t(t)} dW_{2,t} + \frac{d\hat{\sigma}_{FT}[\xi_t + \epsilon \cdot \phi_t]}{d\epsilon} \Big|_{\epsilon=0} dW_{1,t} \quad (\text{B.168})$$

with $\phi_t(u) := \xi_t(u)\tilde{\eta}(u-t)^{-\gamma}$ and where $\hat{\sigma}_{F_T}[\xi]$ is the notation we use to consider the at-the-money implied volatility as a *functional* of the forward variance curve, here evaluated at ξ .

Thus

$$d \ln(S_t) d\hat{\sigma}_{F_T}(t) = \frac{d\hat{\sigma}_{F_T}}{d \ln S} \xi_t(t) dt + \frac{d\hat{\sigma}_{F_T}[\xi_t + \epsilon \cdot \phi_t]}{d\epsilon} \Big|_{\epsilon=0} \sqrt{\xi_t(t)} \rho dt \quad (\text{B.169})$$

and therefore

$$R_T = \frac{1}{\psi(T)} \left(\frac{d\hat{\sigma}_{F_T}}{d \ln S} + \frac{d\hat{\sigma}_{F_T}[\xi_t + \epsilon \cdot \phi_t]}{d\epsilon} \Big|_{\epsilon=0} (\xi_t(t))^{-1/2} \rho \right) \quad (\text{B.170})$$

where we have used that $(d \ln(S_t))^2 = \xi_t(t) dt$ since $d \ln(S_t) = ...dt + \sqrt{\xi_t(t)} dW_{2,t}$.

In our numerical implementation we estimate the above derivatives, including the functional derivative, using a finite difference approximation.

Bibliography

- Andersen, T. G. & Benzoni, L. (2008), ‘Realized Volatility’, *Federal Reserve Bank of Chicago, Working Paper No. 2008-14* .
- Andersen, T. G. & Bollerslev, T. (1997), ‘Intraday periodicity and volatility persistence in financial markets’, *Journal of Empirical Finance*, 4(2):115–158 .
- Barndorff-Nielsen, O. E. & Shephard, N. (2004), ‘Power and bipower variation with stochastic volatility and jumps’, *Journal of Financial Econometrics*, 2(1):1-37 .
- Bayer, C., Friz, P. & Gatheral, J. (2016), ‘Pricing under rough volatility’, *Quantitative Finance*, 16(6):887-904 .
- Bennedsen, M. (2016), ‘Semiparametric inference on the fractal index of gaussian and conditionally gaussian time series data’, *CREATES Research Papers 2016-21, Department of Economics and Business Economics, Aarhus University* .
- Bennedsen, M., Lunde, A. & Pakkanen, M. S. (2017a), ‘Decoupling the short- and long-term behavior of stochastic volatility’, *arXiv:1610.00332v2* .
- Bennedsen, M., Lunde, A. & Pakkanen, M. S. (2017b), ‘Hybrid scheme for brownian semistationary processes’, *Finance and Stochastics*, 21(4):931-965 .
- Bergomi, L. (2009), ‘Smile dynamics IV’, *Risk Magazine*, pages 94-100. Also available at SSRN: <https://ssrn.com/abstract=1520443> .
- Bergomi, L. (2016), *Stochastic Volatility Modelling*, CRC/Chapman & Hall.
- Bergomi, L. & Guyon, J. (2011), ‘The smile in stochastic volatility models’, Available at SSRN: <https://ssrn.com/abstract=1967470> .
- Bernard, C. & Cui, Z. (2010), ‘Pricing timer options’, *Journal of Computational Finance*, Vol. 15, No. 1 .
- Björk, T. (2009), *Arbitrage Theory in Continuous Time*, Oxford University Press.
- Christensen, K., Oomen, R. C. & Podolskij, M. (2014), ‘Fact or friction: Jumps at ultra high frequency’, *Journal of Financial Economics*, 114(3):576–599 .
- Comte, F. & Renault, E. (1998), ‘Long memory in continuous-time stochastic volatility models’, *Mathematical Finance*, 8(4):291-323 .
- Derman, E. & Miller, M. B. (2016), *The Volatility Smile*, Wiley Finance.
- Durrett, R. (1996), *Stochastic calculus: a practical introduction*, volume 6. CRC press.
- Engel, E. & Dreizler, R. M. (2011), *Density Functional Theory*, Springer.

- Euch, O. E., Fukasawa, M., Gatheral, J. & Rosenbaum, M. (2018), ‘Short-term at-the-money asymptotics under stochastic volatility models’, Available at SSRN: <https://ssrn.com/abstract=3111471>
- Fukasawa, M. (2015), ‘Short-time at-the-money skew and rough fractional volatility’, *arXiv:1501.06980v1*.
- Fukasawa, M. (2017), ‘Slides: Hedging and calibration for log-normal rough volatility models’, *Jim Gatheral’s 60th Birthday Conference, Courant Institute, New York University, 13-15th of October 2017*. Available at: <https://wwwf.imperial.ac.uk/~ajacquier/Gatheral60/Slides/Gatheral60%20-%20Fukasawa.pdf>.
- Gatheral, J. (2014), ‘Slides: Fractional volatility models’, *Bloomberg Quantitative Seminar at Bloomberg, New York, June 26th 2014*. Available at: <https://mfe.baruch.cuny.edu/wp-content/uploads/2012/09/FractionalVolatility2014BBQ.pdf>.
- Gatheral, J. & Jacquier, A. (2013), ‘Arbitrage-free SVI volatility surfaces’, *arXiv:1204.0646v4*.
- Gatheral, J., Jaisson, T. & Rosenbaum, M. (2017), ‘Volatility is rough’, *Quantitative Finance* 18(6):933-949, 2018.
- Gawarecki, L. & Mandrekar, V. (2011), *Stochastic Differential Equations in Infinite Dimensions*, Springer.
- Glasserman, P. (2004), *Monte Carlo methods in financial engineering*, Springer.
- Gneiting, T., Ševčíková, H. & Percival, D. B. (2012), ‘Estimators of fractal dimension: Assessing the roughness of time series and spatial data’, *Statistical Science* 2012, 27(2):247-277.
- Hu, D. P. (2003), ‘A remark on non-markov property of a fractional brownian motion’, *Vietnam Journal of Mathematics*, 31(2):237-240.
- Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009), ‘Microstructure noise in the continuous case: The pre-averaging approach’, *Stochastic processes and their applications* 119(7):2249–2276.
- Laakkonen, H. (2007), ‘Exchange rate volatility, macro announcements and the choice of intraday seasonality filtering method’, *Bank of Finland Research Discussion No. 23/2007*. Available at SSRN: <https://ssrn.com/abstract=1032876>.
- Lioui, A. & Poncet, P. (2005), *Dynamic Asset Allocation with Forwards and Futures*, Springer.
- Matteo, T. D. (2007), ‘Multi-scaling in finance’, *Quantitative Finance*, 7(1):21-36.
- McCrickerd, R. & Pakkanen, M. S. (2018), ‘Turbocharging Monte Carlo pricing for the rough Bergomi model’, *arXiv:1708.02563v3*.
- Olver, F. W. J. (1997), *Asymptotics and Special Functions*, A K Peters/CRC Press.
- Oomen, R. C. A. (2006), ‘Realized variance and market microstructure noise’, *Journal of Business and Economic Statistics* 2(24):195–202.
- Sondermann, D. (2006), *Introduction to Stochastic Calculus for Finance*, Springer.